Modelling of the Yield Curve as a New Twist Changed the Plot

A thesis submitted in partial fulfillment of the requirements for the degree MSc in Econometrics and Management Science

G.J. van Dasler

433958

August 10, 2020

Abstract

In this paper we derive two new classes of arbitrage-free specifications within the Dynamic Nelson-Siegel (DNS) environment to alternatively model and forecast the term structure of interest rates. The proposed models combine (i) explicit identification of the theoretical short rate, (ii) the imposition of no-arbitrage dynamics and (iii) bi-directional dependencies between the macroeconomy, a supply factor and the yield curve. Additionally, we introduce a Bayesian algorithm that readily incorporates no-arbitrage dynamics within the Nelson-Siegel environment. Firstly, we show that the inclusion of arbitrage-free conditions results in the desired concavity at the yield curve's long-end for all considered estimation methods. Within our empirical in-sample analysis, we find that the downward pressure at the highest maturities may be too harsh. In contrast, our out-of-sample study shows predictive improvements to emerge from the imposition of no-arbitrage. Secondly, we show that our Bayesian algorithm can beat classical methods within the arbitrage-free Nelson-Siegel environment, although restrictive priors may harm its predictive performance. Lastly, we show that the inclusion of our proposed exogenous factors results in near-consistent improved forecasts at shorter-term and mid-term maturities.

Keywords— Rotated Nelson-Siegel, Arbitrage-Free, Canonical framework, Quantitative Easing, Bayesian Markov Chain Monte Carlo, Scarcity Channel

Supervisor ESE: prof.dr. M. van der Wel

Second assessor ESE: dr. S.O.R. Lönn

Supervisor Rabobank: drs. P.S. Marey

Contents

1 Introduction

Adequate modelling and forecasting of the yield curve are crucial in the fields of trading, pricing of financial derivatives, managing financial risk, allocating assets, structuring debt and policy-making. More general, Estrella & Trubin (2006) and Wright (2006) find the term structure of interest rates to be a leading indicator for the future macroeconomic climate. In order to capture the underlying yield curve dynamics, academics have produced a vast literature with a wide offering of models. Literature traditionally distinguishes two approaches that are successful for complementary reasons. The first approach is theoretically well-motivated, but fails to deliver empirical performance due to identification issues and cumbersome estimation methods. The second approach capitalizes on the yield curve's statistical properties, yet lacks the methodological inclusion of arbitrage-free conditions. Although recent offerings have unified both approaches, it remains infeasible to explicitly identify the short rate within these specifications and to exactly estimate those specifications in a Bayesian manner. In this paper, we overcome both drawbacks and unify exact estimation, theoretical appeal and statistical power in a model that is engaging from a policymaker's perspective.

With this study, we construct a new estimation methodology and several new specifications to capture the yield curve dynamics, both in-sample and out-of-sample. The models we propose combine the statistical power from the Nelson-Siegel environment, the elimination of arbitrage opportunities from traditional models and explicit identification of the theoretical short rate. We provide a greater understanding of the bond market via bi-directional linkages between the yield curve, a newly emerged monetary policy instrument and the macroeconomy. Beyond our derived specifications, we extend on Bayesian estimation methodologies to readily incorporate no-arbitrage dynamics within the Nelson-Siegel framework.

As bonds trade centralized in financial markets, the preclusion of arbitrage opportunities finds empirical motivation and lays the foundation of vast econometric literature. This dates back to early research by Vasicek (1977) and Cox et al. (1985), who assume that a single short-term factor drives the term structure of interest rates. Subsequently, Duffie & Kan (1996) construct a general class of Affine Term Structure Models (ATSMs), although Dai & Singleton (2000) find this framework to suffer from badly behaving likelihoods with nonpositive variances. They compose a canonical representation of admissible ATSMs, i.e. a structured subset of ATSMs with positive variances and well-identified latent factors. Although this class of models shows adequate in-sample performance, it delivers inferior predictive quality compared to random walk forecasts in volatile times (Duffee (2002)). This inability has initiated three developments in econometric literature on which we base our methodology.

The first development comprises returned attention to statistically motivated models. As the term structure of interest rates assumes a wide variety of shapes over time, Nelson & Siegel (1987) combine greater parsimony and flexibility in their three-factor exponential decomposition of the yield curve. The subsequent imposition of time-varying dynamics from Diebold & Li (2006) in Dynamic Nelson-Siegel (DNS) emerged in the eminent workhorse of the industry, i.e. DNS has become the standard for the European Central Bank (Svensson (1995)) and the

Federal Reserve Board (Gürkaynak et al. (2007)). Its popularity stems from several sources. In line with financial theory, DNS beneficially enforces underlying bond prices to approach zero with maturity. Moreover, Diebold & Li (2006) show the emerging latent factors to have meaningful interpretations via a time-varying level, slope and curvature factor, respectively. From a practitioner's perspective, the DNS model additionally offers greater simplicity via its stable and straightforward estimation.

Econometric literature has further iterated on the DNS model with offerings of greater statistical freedom and modifications of the statistical Nelson-Siegel base. The former set finds its origins in the inclusion of time-varying parameters, i.e. Kozicki & Tinsley (2001) allow for a time-varying mean of the process, Koopman et al. (2010) provide support for a time-varying rate of decay and Hautsch & Yang (2012) incorporate stochastic volatility for an in-sample distinction between low-volatile and high-volatile periods. The other set alternated the traditional Nelson-Siegel base, either via greater flexibility or via factor rotation. Per illustration, Svensson (1994) constructs a four-factor model that comprises an additional curvature factor with a second rate of decay. Subsequently, De Rezende & Ferreira (2013) further enlarge its resilience by suggesting a five-factor model with two distinct rates of decay for the slope and curvature factor, respectively. In contrast, the linear transformation from Nyholm (2018) does not allow for greater flexibility as it only changes the interpretation of DNS's level factor to the short rate. His proposed Rotated Nelson-Siegel (RNS) model is statistically equivalent to DNS. Yet, RNS appears more appealing from a policy-making perspective as central banks traditionally seek to accommodate stimuli via manipulation of the short rate.

Despite DNS's empirical success, it misses out on one specific dimension. Björk & Christensen (1999) and Filipović (1999) show its incompetence to be arbitrage-free as DNS does not fit a risk-neutral martingale process. Coroneo et al. (2011) comply with the presence of arbitrage opportunities, yet degrade its impact as they formally reveal DNS and ATSMs to be insignificantly different. Krippner (2015) further formalizes this near-equivalence and demonstrates both theoretical foundation and statistical evidence of DNS to be a low order Taylor approximation of Gaussian ATSMs. Christensen et al. (2011) bridge the gap and impose no-arbitrage dynamics to the DNS model in Arbitrage-Free Nelson-Siegel (AFNS), a slightly restricted version of the canonical representation from Dai & Singleton (2000). On the one hand, the AFNS model circumvents the cumbersome estimation that traditionally accompanies the class of admissible ATSMs. On the other hand, the AFNS model combines the empirical tractability of the DNS model with theoretically appealing arbitrage-free conditions. However, the AFNS model fails to explicitly identify the short rate process with which policymakers traditionally seek to fulfil their mandates. As such, we combine no-arbitrage dynamics from Duffie & Kan (1996) with the factor-interpretation from Nyholm (2018) in Arbitrage-Free Rotated Nelson-Siegel (AFRNS). Our AFRNS model combines the theoretical rigour of ATSMs with the statistical power of DNS while it replaces the level factor with the theoretical short rate. Additionally, we find that the AFRNS model and the AFNS model closely resemble via a linearly interchangeable arbitrage-free parameter.

The second development enriches the class of ATSMs with the inclusion of macroeconomic

factors. Before these developments, literature considered two disjoint philosophies to capture the spot rate dynamics. The macroeconomist approach exploited the role of inflation expectations and real economic activity measures to construct the term structure of interest rates. In contrast, the financial economist approach evaded the involvement of such observable series and resorted to latent factors only. Ang & Piazzesi (2003) combine both strands of literature to describe the collective dynamics between bond prices and approximates for inflation and economic activity within the class of ATSMs. Their extension appears crucial as the observable series explain 85% of variation on the short-end and 40% of variation on the long-end of the yield curve.

Parallel to the proposed extensions for the class of ATSMs, Diebold et al. (2006) find substantial evidence of mutual relations between the macroeconomy and the yield curve in their generalized proposition of the DNS model. De Pooter et al. (2007), Yu & Zivot (2011) and Mönch (2012) exploit these findings in their out-of-sample exercises to find consistent improvement in predictive power. Additionally, Nyholm (2018) link the emerged theoretical short rate with inflation and economic activity in a Taylor rule fashion (e.g. see Taylor (1993)). However, his proposed connections are limited to the evolution of the short rate. Thus, we combine the methods from Diebold et al. (2006) to allow for bi-directional dependencies between the macroeconomy, a later described supply factor and the rotated factors from Nyholm (2018) in Macro-Finance Rotated Nelson-Siegel (MF-RNS). Furthermore, we generalize our AFRNS model in Macro-Finance Arbitrage-Free Rotated Nelson-Siegel (MF-AFRNS) as Ullah (2016) shows the surplus of predictive performance from exogenous factors to prevail with the imposition of no-arbitrage in a Nelson-Siegel environment.

Unusual monetary policy-making has resulted in a climate of tenacious near-zero interest rates, which further harms existing model performance. For example, Steeley (2014) and Trück & Wellmann (2016) show that random walk models consistently outperform more advanced methodologies in the post-crisis era. The effect is two-fold as higher persistence results in further strengthened benchmarks and weaker conventional yield curve models. Andreasen & Meldrum (2013), Bauer et al. (2014) and Christensen & Rudebusch (2015) provide quantitative assessments of the inferior forecast performance from traditional specifications. They show that ATSMs imply negative nominal bond yields and implausible probabilities of negative nominal rates when yields approach zero. These empirical findings conflict with Black (1995), who suggests an option-like approach on the short rate as investors face the option to hold cash rather than to invest in a negative short rate. His theorem states that nominal yields have a non-zero truncation, which imposes the Zero Lower Bound (ZLB) on empirical yields.

The impact of the ZLB is not limited to the term structure of interest rates as it has also been found to diminish existing relations between the term structure and the macroeconomy. Per illustration, the proposed linkages of the short rate to inflation and economic activity from Nyholm (2018) may have declined in explanatory power as Bernanke (2015) finds that the Taylor rule no longer lasts in recent times. Thus, recent advances in literature enrich their yield curve specifications with supply factors, which is in light of the scarcity channel from Modigliani & Sutch (1966) and the preferred habit view from d'Amico et al. (2012). Laubach (2009) already finds that budget deficits and government debt significantly affect the cross-section

of yields before the Great Recession. The subsequent Quantitative Easing (QE) programs have further strengthened the influence of supply levels on the yield curve as Krishnamurty & Vissing-Jorgensen (2011), Ihrig et al. (2012) and Greenwood & Vayanos (2014) find significant downward pressure from Large Scale Asset Purchases (LSAPs) on the long-end of the yield curve. Additionally, Li & Wei (2012) explicitly involve supply levels of long-term treasuries and Mortgage-Backed Securities (MBS) in the framework from Ang & Piazzesi (2003). They find a one percentage point depression on the yield curve's long-end from LSAP2. Hence, we account for the alternated interest rate climate as we complement the inclusion of approximates for inflation and economic activity with the Marketable Fraction (MF) of long-term treasuries.

The third development involves the greater interest of recent literature in Bayesian estimation methodologies, mostly due to their statistical advantages. Firstly, Bayesian analysis circumvents the issue of problematic local maxima (Kim & Wright (2005)), i.e. points with similar likelihoods but vastly different interpretations of the parameter coefficients. Illustratively, Ang et al. (2007) adopt a Bayesian framework to cope with burdensome estimation arising from nonlinearity in the deep parameters of ATSMs. Secondly, Bayesian inference supports the incorporation of basic economic intuition via sensible prior densities (Diebold et al. (2008)). For instance, Chib & Ergashev (2009) incorporate the belief of a positive term premium to support the otherwise ill-determined parameter estimation of ATSMs, considerably enhancing out-of-sample performance. Thirdly, Bayesian analysis demonstrates robustness to the initialization routine, which becomes increasingly burdensome when the amount of parameters involved becomes large. For example, Christensen et al. (2011) stress classical estimation of the canonical framework from Dai & Singleton (2000) to be both onerous and time-consuming and Ullah (2016) shows several preliminary costly initialization routines to be essential for likelihood convergence to reasonable estimates. Bayesian analysis, however, needs no preliminary procedures. Lastly, Bayesian methodologies beneficially come with quantitative assessment of model-, parameter-, and forecast uncertainty. Recently, Pericoli & Taboga (2018) implement an exact Bayesian framework to provide a quantitative assessment of the divergence between institutional policy-making for different monetary unions.

These Bayesian advances are not limited to the class of ATSMs as they extend to other, statistically-inspired yield curve specifications. For example, Mönch (2012) analyzes the predictive content of level, slope and curvature as he models the joint dynamics of macroeconomic variables and interest data. Diebold et al. (2008) estimate the DNS model using Bayesian Markov Chain Monte Carlo (MCMC) techniques, where Hautsch & Ou (2008), Caldeira et al. (2010) and Hautsch & Yang (2012) subsequently iterate via the allowance for time-varying parameters and stochastic volatility in the DNS model. Laurini & Caldeira (2016) provide even greater flexibility and construct a Bayesian sampler to combine both time-varying parameters and the inclusion of exogenous factors. More recently, Çakmaklı (2020) seeks to model the post-crisis low-volatility period with the imposition of a Dirichlet process prior.

Instead, we exploit the unrestricted real-world dynamics to reversely tie our Nelson-Siegel specifications to the arbitrage-free framework from Duffie & Kan (1996) within a Bayesian estimation method. To our knowledge, no advances in Bayesian methodologies have yet

been made to readily incorporate no-arbitrage dynamics within the Nelson-Siegel environment. Traditionally, Christensen et al. (2011) and Ullah (2016) opt for maximum likelihood estimation via the Nelder-Mead simpled method (e.g. see Singer & Nelder (2009)). However, we follow Diebold & Rudebusch (2013) as they debate the appealing properties of a Bayesian approach to make DNS arbitrage-free as it hypothetically provides a natural shrinkage direction.

We consider end-of-month U.S. zero-coupon bond yields from July 2003 to February 2020 to conduct an empirical exercise on our proposed extensions. In order to isolate each of the modelled aspects, we address the following research questions:

1. Does the allowance for correlated factor dynamics provide additional explanatory power in our empirical framework?

2. How does the prohibition of arbitrage opportunities affect the fit and the predictive performance of our considered specifications?

3. How does the factor rotation from Nyholm (2018) affect the in-sample and out-of-sample performance, when compared to the traditional Nelson-Siegel base?

4. Does our Bayesian estimation method provide greater explanatory power when compared to classical estimation methods?

5. Does the inclusion of our proposed exogenous regressors grant extra performance in both analyses, when compared to their yields-only counterparts?

We solely investigate the fourth research question for yields-only models (e.g. specifications that do not implement observable series other than yields). More specific, Ullah (2016) shows that the classical estimation of models that combine exogenous factors and no-arbitrage within the Nelson-Siegel environment requires costly initialization routines via simulation. We consider these procedures to be beyond the scope of this paper. Besides, we find that comparison of our proposed methodology with the classical estimation method for yields-only specifications already provides sufficient evidence to answer the fourth question.

Within our in-sample exercise, we find that the mean fit of the term structure is relatively robust to all modelled aspects presented in the research questions. Specifically, we find that the deviation in average fit maximally amounts 1.5 basis points for all modifications considered. When we divide the in-sample fit per maturity, some deviations arise. First, we find consistent overestimation of the yield curve's short-end, which indicates the inadequacy of our proposed modifications to model the near-zero yields for lower maturities during the ZLB period (Bauer et al. (2014)). This particularly holds for the arbitrage-free models from the second research question. On the long-end, however, we find that the statistical specifications cannot capture the desired empirical concavity. The arbitrage-free specifications can account for the implied flattening, yet, we find them to exert too much downward pressure for higher maturities. This overestimation prevails with both Nelson-Siegel bases and estimation procedures from the third and fourth research questions, respectively. This indicates that the arbitrage-free restrictions may be too harsh to model the long-end during the ZLB period. Although the

inclusion of exogenous factors from the fifth research question does not yield improvements to the in-sample fit, we find it to offer a greater understanding of the mutual dependencies between the macroeconomy, the monetary policy stance and the term structure of interest rates.

Our out-of-sample analysis shows more diverging results for each of the modelled aspects. With regards to the first research question, we find that correlated factor specifications offer higher predictive performance in forecasting the short-end of the term structure. This performance slowly dissipates towards the long-end of the yield curve, where we find better performance with independent factor models. We obtain conflicting evidence on whether the imposition of no-arbitrage dynamics from the second research question offers better forecasts or not. For the AFNS model, we find a near-uniform improvement compared to its no-arbitrage counterpart. This is further strengthened by our proposed estimation method from the fourth research question, i.e. we find that our Bayesian estimation methodology for the AFNS model offers increased performance in a predictive setting when compared to the classical method. Contrarily, the AFRNS model can offer greater predictive performance as opposed to its statistical peer, yet the imposed priors from our Bayesian estimation methodology appear too binding for this model. For the statistical models, we observe that the predictive strength of our Bayesian approach slowly increases towards the long-end. The near-equivalence between both statistical bases from the third research question largely prevails with our out-of-sample exercise. Regarding the fifth research question, we find forecasting improvements for yields with shorter-term and mid-term maturities to emerge from the inclusion of our proposed observable factors.

We organize the remainder of this paper as follows. In Section 2, we depart from the traditional Nelson-Siegel environment and elaborate on our proposed extensions. Section 3 describes the implemented estimation procedures for all models considered. Next, Section 4 specifies the data used in the empirical exercises from Section 5. We conclude our findings in Section 6 and finalize with some suggestions for further research in Section 7.

2 Econometric Methods & Techniques

We elaborate in Section 2.1 on the statistical workhorse of the industry, the Dynamic Nelson-Siegel (DNS) model. We explicitly identify the theoretical short rate as the first latent factor while we preserve the statistical power of the DNS model with the Rotated Nelson-Siegel (RNS) model from Nyholm (2018) in Section 2.2. In Section 2.3, we introduce a new specification as we impose the RNS model to the arbitrage-free structure from Duffie & Kan (1996). Alternatively, we combine the factor rotation from Nyholm (2018) with modelled bi-directional dependencies between exogenous factors and the latent states via Diebold et al. (2006) in Section 2.4. Lastly, we unify the proposed extensions from Sections 2.3 and 2.4 within one new specification in Section 2.5.

2.1 Dynamic Nelson-Siegel

As the term structure of interest rates assumes a wide variety of shapes over time, Nelson & Siegel (1987) introduce greater parsimony and flexibility via Laguerre polynomial decomposition of the forward rate $f_t(\tau_i)$. If we define $P_t(\tau_i)$ as the price of a zero-coupon bond expiring at time $t+\tau_i$ for $i = 1, \dots, N$ and $y_t(\tau_i)$ as the annualized rate of return resulting from continuous compounding such that $P_t(\tau_i) = e^{-y_t(\tau_i)\tau_i}$, differential equation $y_t(\tau_i) = \frac{1}{\tau_i}\int_0^{\tau_i} f_t(u)\mathrm{d}u$ implies the yield curve at time t to be defined as

$$y_t(\tau_i) = f_{1,t} + f_{2,t}\left[\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right] + f_{3,t}\left[\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right], \qquad i = 1, 2, \dots, N, \tag{1}$$

where λ governs the exponential decay rate and $f_{j,t}$ refers to the j-th latent factor at time t with $j = 1, 2, 3$. Large (small) values of λ correspond to fast (slow) decay with maturity. Apart from the statistical advantages the Nelson-Siegel functional form provides, it beneficially implements empirical restrictions from financial theory as $P_t(0) = 1$ and $\lim_{\tau_i\to\infty} P_t(\tau_i) = 0\ \forall t$. The introduction of time-varying dynamics from Diebold & Li (2006) to capture bond market movements over time in the Dynamic Nelson-Siegel (DNS) model largely increased its popularity as it has become the yardstick among practitioners (Svensson (1995); Curves (2005); Gürkaynak et al. (2007)).

Analogue to Principal Component Analysis (PCA), Diebold & Li (2006) suggest to interpret the time-varying latent factors $\{f_{1,t}, f_{2,t}, f_{3,t}\}_{t=1}^{T}$ as level, slope and curvature, respectively. Literature considers $f_{1,t}$ a long-term factor as $\lim_{\tau_i\to\infty} y_t(\tau_i) = f_{1,t}$. A shock thereof equally affects the entire cross-section of modelled yields due to its unity loading. Frankel & Lown (1994) interpret $f_{2,t} = \lim_{\tau_i\to 0} y_t(\tau_i) - \lim_{\tau_i\to\infty} y_t(\tau_i)$ as a negative slope whose loading starts at one and monotonically approaches zero. Sudden changes on the factor highly affect short-term yields and induce variations in the spreads. Rudebusch & Wu (2008) relate the first factor to market expectations on the policymaker's medium-term target inflation. Additionally, they show the slope factor to comply with cyclical responses of the central bank's short rate manipulation to fulfil its mandate. The third latent factor portrays a medium-term factor as its loading bulks roughly around mid-term maturities depending on λ, i.e. it increases for short-term maturities but influence fades towards the long-end of the term structure. Mönch (2012) shows unexpected increments of the third factor to come before a flattening yield curve and possible economic slowdown (Wright (2006)). Besides, Figure 1 displays the behaviour of DNS's factor loadings as a function of maturity, i.e. we observe the unity loading on the first factor, the gradual decay on the second factor loading and the bell-shaped pattern on the third factor loading.

We follow Diebold et al. (2006) who suggest to rewrite the model in state space with further eased single-step estimation, i.e.

$$\boldsymbol{y_t} = \boldsymbol{B_\lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \tag{2}$$

$$\boldsymbol{f_t} = (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{f}_{t-1} + \boldsymbol{\eta_t}, \qquad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, 2, \dots, T, \tag{3}$$

where Equations (2) and (3) refer to the measurement equation and state equation, respectively.

Equation (2) reformulates Equation (1) to a multivariate setting, i.e. it bundles N observed yields and three factors at time t in $\boldsymbol{y_t}$ and $\boldsymbol{f_t}$. The covariance matrix on the measurement equation is assumed to be diagonal, i.e. $\boldsymbol{H} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. Explicit completion of the factor loadings yields

$$\boldsymbol{B}^*_{\boldsymbol{\lambda}} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix}, \tag{4}$$

as an $N \times 3$ matrix. Equation (3) resembles either an AR(1) or a VAR(1) process on the latent factors with coefficients captured in state dynamics matrix $\boldsymbol{F}$ (3×3) and covariance matrix $\boldsymbol{Q}$ (3×3). An AR(1) process imposes the latent factors to be independent via a diagonal structure on both $\boldsymbol{F}$ and $\boldsymbol{Q}$ (Diebold & Li (2006)), whereas a VAR(1) process allows for correlated dynamics via a fully-flexible $\boldsymbol{F}$ and Cholesky-decomposable $\boldsymbol{Q}$ (Diebold et al. (2006)).

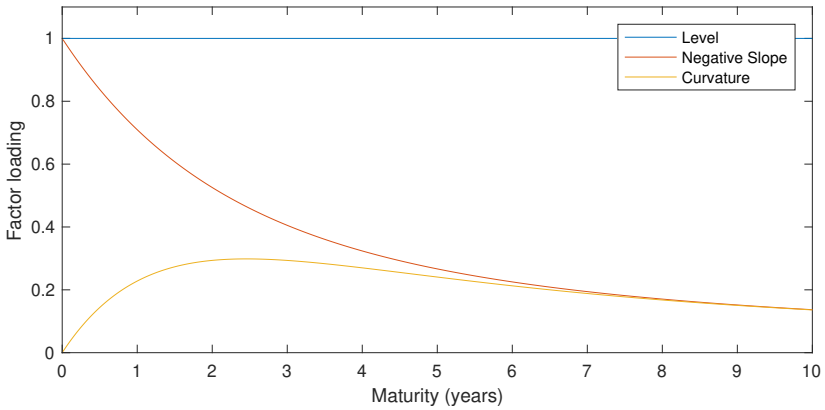

Figure 1: Factor loadings acquired from the Dynamic Nelson-Siegel (DNS) model for each entry provided in the factor loading matrix of Equation (4). The rate of decay λ is fixed to 0.7308, which maximizes the factor loading on the yields with 30 months to maturity (Diebold & Li (2006)).

2.2 Rotated Nelson-Siegel

The functional form of Nelson & Siegel (1987) does not directly parametrize the overnight rate as it corresponds to the sum of the first two latent factors by definition, i.e. $\lim_{\tau_i \to 0} y_t(\tau_i) = f_{1,t} + f_{2,t}$. From a regulatory view, policymakers manipulate the short rate to reach their inflation targets and to accommodate economic growth. Hence, explicit identification of the overnight rate enables the interrogation of joint dynamics between the short rate and macroeconomic factors in a Taylor Rule type fashion and possibly surpluses the explanatory power of DNS. Besides, a plain expression on the short rate process has become increasingly important since its purpose as traditional monetary policy instrument cuts no more slack (Rudebusch et al. (2018)) and its predictive use as liftoff indicator has gained attention (Bauer et al. (2016); Feroli et al. (2017)).

To overcome this obstacle, Nyholm (2018) augments the DNS model with rotation matrix $\boldsymbol{A}$ after which the factor-interpretation of interest emerges in the Rotated Nelson-Siegel (RNS)

model. Explicit completion of $\boldsymbol{A}$ where the theoretical short rate emerges on the first latent factor yields

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

which alternates the parameter description from [Level, Negative Slope, Curvature] to [Short Rate, Positive Slope, Curvature].[1] Hence, the yield curve expression of the RNS model at time t resolves to

$$y_t(\tau_i) = \beta_{1,t} + \beta_{2,t}\left[1 - \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}\right] + \beta_{3,t}\left[\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right], \qquad i = 1, 2, \ldots, N, \tag{6}$$

where the proposed transition rigorously shifts the first factor from being long-term to short-term, i.e. $\lim_{\tau_i \to 0} y_t(\tau_i) = f_{1,t} + f_{2,t} = \beta_{1,t}$. The preservation of its unity loading inversely constructs each modelled yield to be loaded on the theoretical short-rate instead of the level factor. Analogue to the level-interpretation of the DNS model, a shock on the short rate equally affects the entire cross-section of yields (Diebold & Rudebusch (2013)). The second factor switches sign compared to the DNS model as Frankel & Lown (1994) show $f_{2,t} = \lim_{\tau_i \to 0} y_t(\tau_i) - \lim_{\tau_i \to \infty} y_t(\tau_i) = -\beta_{2,t}$. Hence, the second factor refers to the theoretical spread and shows inverse decay as the influence of the slope gradually increases towards the long-end of the curve. Curvature-akin interpretation persists on the third latent factor (Nyholm (2018)). Additionally, Figure 2 displays the evolution with maturity of the factor loadings from the RNS model. Comparison of Figures 1 and 2 demonstrates that the sole deviation between both bases lies in the horizontally mirrored pattern of the second factor loading.

To further ease notation and estimation, Nyholm (2018) suggests to rewrite the static composition from Equation (6) in state space as

$$\boldsymbol{y_t} = \boldsymbol{B^*_\lambda}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \tag{7}$$

$$\boldsymbol{\beta_t} = (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta_t}, \qquad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, 2, \ldots, T, \tag{8}$$

where Equations (7) and (8) refer to the measurement equation and state equation of the RNS model, respectively. Both show great resemblance to the state space formulation of the DNS model in Equations (2) and (3), i.e. both formulations are dimensional equivalent, although the completion of the parameters, factors and disturbances is different due to the factor rotation.[2]

[1] We implement the short rate as a theoretical quantity throughout our paper, i.e. the first latent factor to emerge has maturity $\tau = 0$, as the resulting completion of rotation matrix $\boldsymbol{A}$ yields greater simplicity to subsequently introduce no-arbitrage dynamics. However, Nyholm (2018) generalizes the completion of transformation matrix $\boldsymbol{A}$ to define the first latent factors as an empirical approximate of the short rate that f.e. has maturity $\tau = 3$ or $\tau = 6$ in months.

[2] The completion of the parameters and disturbances in Equations (2) and (3) differ due to the factor augmentation from Nyholm (2018). For readability purposes we only introduce new notation for the states and the factor loadings as we wish to directly compare these two aspects between both specifications.

Completion of the factor loading matrix from the RNS model resolves to

$$\boldsymbol{B}^*_{\boldsymbol{\lambda}} = \begin{bmatrix} 1 & 1 - \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & 1 - \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & 1 - \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{bmatrix}, \tag{9}$$

where the entries in the second column differ from the traditional Nelson-Siegel base in Equation (4). As $\boldsymbol{A}$ is involutory and $\boldsymbol{B}^*_{\boldsymbol{\lambda}}\boldsymbol{\beta}_t = \boldsymbol{B}_{\boldsymbol{\lambda}}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{f}_t = \boldsymbol{B}_{\boldsymbol{\lambda}}\boldsymbol{f}_t$, both the DNS model and the RNS model are observational equivalent with identical statistical characteristics in their measurement equations. However, this interchangeability no longer prevails with the incorporation of exogenous variables. As the first factor embodies an alternative statistical interpretation in both specifications, an improvement of one variant over the other resolves to an empirical question with no statistical consensus in case we implement exogenous regressors. For example, if the exogenous factors are closely related to the theoretical short rate (level), we expect the RNS (DNS) model to posses greater explanatory power.

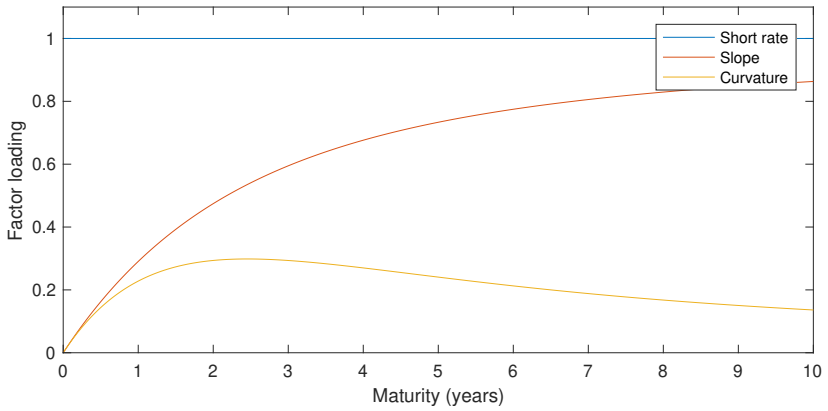

Figure 2: Factor loadings acquired from the Rotated Nelson-Siegel (RNS) model for each entry provided in the factor loading matrix in Equation (9). The rate of decay λ is fixed to 0.7308, which maximizes the factor loading on the yields with 30 months to maturity (Diebold & Li (2006)).

2.3 Arbitrage-Free Rotated Nelson-Siegel

We depart from the RNS model and combine the absence of arbitrage opportunities with direct parametrization of the short rate in Arbitrage-Free Rotated Nelson-Siegel (AFRNS). From one perspective, the AFRNS model circumvents troublesome estimation that traditionally accompanies the class of ATSMs (Duffee (2002); Kim & Wright (2005); Hamilton & Wu (2012)). Alternatively, the AFRNS model methodologically surpluses the empirically successful but theoretically lacking RNS model as the latter misses out on the no-arbitrage dimension (Duffee (2002); Kim & Wright (2005); Hamilton & Wu (2012)) while explicit identification of the theoretical short rate persists with the imposition of no-arbitrage.

The existence of the AFRNS model originates from two conditions. Firstly, the existence of a riskless profit hindering Dynamic Nelson-Siegel specification stems from the work of

Trolle & Schwartz (2009). They show the HJM-framework to be readily represented by a multi-dimensional Markov process with time-invariant volatility structures (e.g. see Heath et al. (1992) for an explanation in greater detail). As the DNS model satisfies this requirement, there exists an arbitrage-free three-dimensional HJM model with the traditional Nelson & Siegel (1987) base. Secondly, the statistical equivalence of RNS and DNS philosophically ensures the same principle to hold for a no-arbitrage variant of the RNS model (Nyholm (2018)).

The derivation of the AFRNS model is analogue to the derivation of Arbitrage-Free Rotated Nelson-Siegel (AFNS) from Christensen et al. (2011) and commences with the imposition of the RNS model on the continuous-time no-arbitrage framework from Duffie & Kan (1996). Following Christensen et al. (2011) we assume a three-dimensional state variable $\boldsymbol{X_t}$ to be a Gaussian Ornstein-Uhlenbeck process under the risk-neutral $\mathbb{Q}$-measure with Stochastic Differential Equation (SDE)

$$\mathrm{d}\boldsymbol{X_t} = \boldsymbol{K_t^{\mathbb{Q}}}[\boldsymbol{\theta_t^{\mathbb{Q}}} - \boldsymbol{X}_t]\mathrm{d}t + \boldsymbol{\Sigma_t D_t(X_t)}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}}, \tag{10}$$

where $\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}}$ (3×1) is a multivariate standard Brownian motion under the risk-neutral measure of which filtration $\{\mathcal{I}_t : t \geq 0\}$ is compliant with the usual conditions from Williams (1991) (e.g. we refer to his study for an explanation in greater detail). Both $\boldsymbol{\theta_t^{\mathbb{Q}}}$ (3×1) and $\boldsymbol{K_t^{\mathbb{Q}}}$ (3×3) are bounded continuous functions, which correspond to the time-varying drift and dynamics of the process under the $\mathbb{Q}$-measure respectively. The risk-neutral volatility of the process consists of two components, i.e. a bounded continuous matrix $\boldsymbol{\Sigma_t}$ (3×3) and a diagonal matrix $\boldsymbol{D_t(X_t)}$ (3×3) with elements $\sqrt{\gamma_t^j + (\boldsymbol{\delta_t^j})'\boldsymbol{X_t}}$, where scalar γ_t^j and vector $\boldsymbol{\delta_t^j}$ (3×1) are bounded continuous functions for $j = 1, 2, 3$. The instantaneous risk-free rate coincides with an affine function of the unobserved state variables as

$$r_t = \rho_{0,t} + \boldsymbol{\rho_{1,t}'}\boldsymbol{X_t}, \tag{11}$$

which fits the RNS model via $\rho_{0,t} = 0$ and $\boldsymbol{\rho_{1,t}} = [1, 0, 0]'$ $\forall t$ as $\lim_{\tau_i \to 0} y_t(\tau_i) = \beta_{1,t}$ (e.g. see Section 2.2). Duffie & Kan (1996) prove zero-coupon bond prices to be exponential functions of the underlying latent factors under the risk-neutral measure, i.e.

$$P(t,T) = \mathbb{E}_t^{\mathbb{Q}}\Big[e^{\left(-\int_t^T r_u \mathrm{d}u\right)}\Big] = e^{A(t,T) + \boldsymbol{B(t,T)}'\boldsymbol{X_t}}, \tag{12}$$

where $A(t,T)$ and $\boldsymbol{B(t,T)}$ form time-invariant solutions to the following system of Riccati Ordinary Differential Equations (ODEs)

$$\frac{\mathrm{d}A(t,T)}{\mathrm{d}t} = \rho_0 - \boldsymbol{B(t,T)}'(\boldsymbol{K^{\mathbb{Q}}})\boldsymbol{\theta^{\mathbb{Q}}} - \frac{1}{2}\sum_{j=1}^{3}\Big(\boldsymbol{\Sigma}'\boldsymbol{B(t,T)}\boldsymbol{B(t,T)}'\boldsymbol{\Sigma}\Big)_{j,j}\gamma^j, \tag{13}$$

$$\frac{\mathrm{d}\boldsymbol{B(t,T)}}{\mathrm{d}t} = \boldsymbol{\rho_1} + (\boldsymbol{K^{\mathbb{Q}}})'\boldsymbol{B(t,T)} - \frac{1}{2}\sum_{j=1}^{3}\Big(\boldsymbol{\Sigma}'\boldsymbol{B(t,T)}\boldsymbol{B(t,T)}'\boldsymbol{\Sigma}\Big)_{j,j}(\boldsymbol{\delta^j}). \tag{14}$$

The solutions to the ODEs in Equations (13) and (14) comply with boundary conditions

$A(T,T) = 0$ and $\boldsymbol{B(T,T) = 0}$ as $P(T,T) = 0$.[3] Following Christensen et al. (2011) we suppress the time-dependence on the parameters in Equation (10). Moreover, we specify the risk-neutral evolution on the state variables to be homoskedastic with $\boldsymbol{D_t(X_t)} = \boldsymbol{I_3}$ $\forall t$. Risk-neutral pricing of zero-coupon bonds according to Equation (12) implies closed-form computation of the underlying yields via

$$y(t,T) = -\frac{\log P(t,T)}{T-t} = -\frac{A(t,T)}{T-t} - \frac{\boldsymbol{B(t,T)'}}{T-t}\boldsymbol{X_t}, \tag{15}$$

which roughly follows the structure in Equation (6). An approximate match between the three-factor affine structure from Duffie & Kan (1996) and the augmented base from Nyholm (2018) yields explicit solutions to the Riccati ODEs as

$$B^1(t,T) = -(T-t), \tag{16}$$

$$B^2(t,T) = -(T-t) + \frac{1-e^{-\lambda(T-t)}}{\lambda}, \tag{17}$$

$$B^3(t,T) = (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda}, \tag{18}$$

which we formalize with Equations (47), (48) and (49) in Appendix A.1. We obtain a formal solution for $A(t,T)$ via

$$A(t,T) = \sum_{j=2}^{3}\left[(\boldsymbol{K}^{\mathbb{Q}}\boldsymbol{\theta}^{\mathbb{Q}})_j \int_t^T B^j(s,T)\mathrm{d}s\right] + \frac{1}{2}\sum_{j=1}^{3}\int_t^T \left(\boldsymbol{\Sigma}'\boldsymbol{B(s,T)B(s,T)'}\boldsymbol{\Sigma}\right)_{j,j}\mathrm{d}s. \tag{19}$$

Hence, we obtain a near-exact one-to-one mapping of the three-factor affine model from Duffie & Kan (1996) to the factor loadings of the RNS model, yet the inclusion of an extra yield-adjustment term $-\frac{A(t,T)}{T-t}$ arises from Jensen's inequality (Christensen et al. (2011)).

Although we provided some intuition on the unification of the RNS model with the arbitrage-free structure from Duffie & Kan (1996), we elaborate on the construction of the AFRNS specification in Appendix A. Appendix A.1 clarifies the risk-neutral state dynamics of the AFRNS model for which we provide a formal proof in Appendix A.2, i.e. we inversely show the suggested solutions to match the ODEs and boundary conditions. To conclude our proposition of the AFRNS model, we construct a simplified expression for the proposed bias correction in Equation (19) and show the yield-adjustments of the AFRNS model and the AFNS model to be linearly interchangeable in Appendix A.3.

In Appendix A.1 we show that the AFRNS model fits the risk-neutral state diffusion of

[3] We opt for alternative notation here of $P(t,T)$ over $P_t(\tau)$ with $T = t+\tau$ to conform with the documentation used by Christensen et al. (2011). The notation is interchangeable, but further eases visual inspection of the similarities between the derivations of the AFRNS model and the AFNS model.

Equation (10) via

$$\begin{pmatrix} \mathrm{d}X_t^1 \\ \mathrm{d}X_t^2 \\ \mathrm{d}X_t^3 \end{pmatrix} = \begin{bmatrix} 0 & -\lambda & -\lambda \\ 0 & \lambda & \lambda \\ 0 & 0 & \lambda \end{bmatrix} \left[\begin{pmatrix} \theta_1^{\mathbb{Q}} \\ \theta_2^{\mathbb{Q}} \\ \theta_3^{\mathbb{Q}} \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] \mathrm{d}t + \boldsymbol{\Sigma} \begin{pmatrix} \mathrm{d}W_{1,t}^{\mathbb{Q}} \\ \mathrm{d}W_{2,t}^{\mathbb{Q}} \\ \mathrm{d}W_{3,t}^{\mathbb{Q}} \end{pmatrix}, \quad \lambda > 0, \tag{20}$$

which has several mathematical implications. Firstly, the explicit filling of $\rho_{0,t}$ and $\boldsymbol{\rho_{1,t}}$ in Equation (11) ensures explicit identification of the first latent factor as risk-free rate to prevail with the imposition of no-arbitrage. In line with its role in the AFNS model, curvature's function as a time-varying mean for the slope factor remains with the factor rotation from Nyholm (2018) under the risk-neutral measure (e.g. see the placement of λ as the third element of the second row in Equation (20)). However, its role expands as both slope and curvature form stochastic additions to the risk-neutral mean of the short rate factor. Secondly, the short rate factor has a unit root under the $\mathbb{Q}$-measure as the first diagonal element of $\boldsymbol{K}^{\mathbb{Q}}$ in Equation (20) is zero. This second implication conforms with empirical evidence as Cieslak (2018) shows the overnight rate to be nonstationary in recent times. Thirdly, the imposition of no-arbitrage in the AFRNS model statistically limits the risk-neutral state diffusion with a time-invariant unconditional mean $\boldsymbol{\theta}$, a time-invariant risk-neutral volatility matrix $\boldsymbol{\Sigma}$ and a time-invariant rate of decay λ. The first restriction directly impacts the statistical freedom of the AFRNS model as it becomes incompatible with the greater explanatory performance arising from shifting endpoints (Orphanides & Wei (2012); Van Dijk et al. (2014); Altavilla, Giacomini, & Ragusa (2017)). The second constraint implies a joint incorporation of no-arbitrage and unspanned stochastic volatility via the methods from Kim et al. (1998) to be infeasible. The third limitation prohibits the implementation of the extensions with a time-varying rate of decay from Koopman et al. (2010) in a no-arbitrage fashion.

In line with the continuous-time representation of the AFNS model from Christensen et al. (2011), we derive the continuous-time relationship between the risk-neutral and real-world dynamics of the AFRNS model via measure change

$$\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}} = \mathrm{d}\boldsymbol{W_t^{\mathbb{P}}} + \boldsymbol{\Gamma_t}\mathrm{d}t, \tag{21}$$

where $\boldsymbol{\Gamma_t}$ refers to the risk premium, $\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}}$ coincides with the risk-neutral Brownian motion from Equation (10) and $\mathrm{d}\boldsymbol{W_t^{\mathbb{P}}}$ (3×1) corresponds to a standard Brownian motion under the $\mathbb{P}$-measure. Following Duffee (2002) we construct the risk premium to be a linear function of the underlying latent state variables as

$$\boldsymbol{\Gamma_t} = \begin{pmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \gamma_3^0 \end{pmatrix} + \begin{bmatrix} \gamma_{11}^1 & \gamma_{12}^1 & \gamma_{13}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 \\ \gamma_{31}^1 & \gamma_{32}^1 & \gamma_{33}^1 \end{bmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix}, \tag{22}$$

Table 1: AFRNS parameter restrictions on $A_0(3)$ class of models.

AFRNS	$\boldsymbol{\rho_0^Y}$, $\boldsymbol{\rho_1^Y}$	$\boldsymbol{K^{Y,\mathbb{Q}}}$	$\boldsymbol{K^{Y,\mathbb{P}}}$	$\boldsymbol{\theta^{Y,\mathbb{P}}}$	**Amount of Restrictions**
Independent factors	$\rho_0^Y = \rho_{1,2}^Y = \rho_{1,3}^Y = 0$	$\kappa_{11}^{Y,\mathbb{Q}} = 0$ $\kappa_{22}^{Y,\mathbb{Q}} = \kappa_{33}^{Y,\mathbb{Q}}$ $\kappa_{23}^{Y,\mathbb{Q}} = \kappa_{22}^{Y,\mathbb{Q}}\left(\frac{\kappa_{13}^{Y,\mathbb{Q}}}{\kappa_{12}^{Y,\mathbb{Q}}}\right)$	Diagonal	No restrictions	12
Correlated factors	$\rho_0^Y = 0$	$\kappa_{11}^{Y,\mathbb{Q}} = 0$ $\kappa_{22}^{Y,\mathbb{Q}} = \kappa_{33}^{Y,\mathbb{Q}}$	No restrictions	No restrictions	3

Note: This table provides the supplemental required restrictions on the canonical $A_0(3)$ model to correctly identify both independent factor- and correlated factor AFRNS. The same theoretical restrictions apply to the MF-AFRNS specifications. Following Christensen et al. (2011) we provide an extensive derivation of these restrictions in Appendix B.

which extends the affine dynamics from Equation (20) to the $\mathbb{P}$-measure, i.e.

$$\begin{aligned} \mathrm{d}\boldsymbol{X_t^{\mathbb{P}}} &= \boldsymbol{K^{\mathbb{P}}}[\boldsymbol{\theta^{\mathbb{P}}} - \boldsymbol{X_t}] + \boldsymbol{\Sigma}\mathrm{d}\boldsymbol{W_t^{\mathbb{P}}} \\ &= \begin{bmatrix} \kappa_{11}^{\mathbb{P}} & \kappa_{12}^{\mathbb{P}} & \kappa_{13}^{\mathbb{P}} \\ \kappa_{21}^{\mathbb{P}} & \kappa_{22}^{\mathbb{P}} & \kappa_{23}^{\mathbb{P}} \\ \kappa_{31}^{\mathbb{P}} & \kappa_{32}^{\mathbb{P}} & \kappa_{33}^{\mathbb{P}} \end{bmatrix} \left[\begin{pmatrix} \theta_1^{\mathbb{P}} \\ \theta_2^{\mathbb{P}} \\ \theta_3^{\mathbb{P}} \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] \mathrm{d}t + \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{32} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{P}} \\ \mathrm{d}W_t^{2,\mathbb{P}} \\ \mathrm{d}W_t^{3,\mathbb{P}} \end{pmatrix}. \end{aligned} \tag{23}$$

The greater flexibility of $\boldsymbol{\Gamma_t}$ in Equation (22) motivates the exploitation of the inverse expressions from Fisher & Gilles (1996) (e.g. see Section 3) as the unconditional mean vector $\boldsymbol{\theta^{\mathbb{P}}}$ and the mean-reversion matrix $\boldsymbol{K^{\mathbb{P}}}$ are allowed to move freely under the $\mathbb{P}$-measure.

Two specific sets of models from Christensen et al. (2011) emerge from Equation (23), i.e. the class of independent factor models and the class correlated factor models. The former provides greater parsimony as it restricts off-diagonal elements of both $\boldsymbol{K^{\mathbb{P}}}$ and $\boldsymbol{\Sigma}$ to zero, whereas the latter offers larger resilience with unrestricted cross-sectional state dynamics and mutual interaction between shocks on the state variables via a lower-triangular matrix $\boldsymbol{\Sigma}$.

Both variants of AFRNS require several additional restrictions to conform with the canonical $A_0(3)$ class of no-arbitrage specifications from Dai & Singleton (2000), i.e. a generic class of models with a three-factor latent state diffusion, zero square-root volatility processes and no-arbitrage. Per illustration, the acquired latent factors may rotate in absence as the probability distribution of bond yields remains unchanged (Duffee (2011)). We elaborate on the compulsory restrictions made by Dai & Singleton (2000) in Appendix B and show these to be incapable of identifying our considered AFRNS specifications. Consequentially, we provide supplemental restrictions to correctly identify the system as we prohibit affine transformations to yield equivalent representations. To clarify, we ensure an arbitrary three-factor state diffusion $\boldsymbol{Y_t}$ to be unique as we prohibit a linear transformation $\mathcal{T}_C$: $\boldsymbol{Z_t} = \boldsymbol{C}\boldsymbol{Y_t} + \boldsymbol{\epsilon}$ with non-singular square matrix $\boldsymbol{C}$ (3×3) and vector of constants $\boldsymbol{\epsilon}$ (3×3) to result in an identical representation of the same model.

Table 1 summarizes the supplemental parameter restrictions from Appendix B for both the independent factor AFRNS model and the correlated factor AFRNS model. In line with the notation from Duffie & Kan (1996), $\boldsymbol{\rho_0^Y}$ and $\boldsymbol{\rho_1^Y}$ refer to the linear coefficients used to construct the short rate from the underlying factors $\boldsymbol{Y_t}$, $\boldsymbol{K^{Y,\mathbb{Q}}}$ and $\boldsymbol{K^{Y,\mathbb{P}}}$ correspond to the risk-neutral dynamics and the real-world dynamics of $\boldsymbol{Y_t}$, respectively, and $\boldsymbol{\theta^{Y,\mathbb{P}}}$ coincides with the drift of

$\boldsymbol{Y_t}$ under the $\mathbb{P}$-measure. In line with the compulsory restrictions on the AFNS model, jointly restricting $\rho_0^Y = \kappa_{11}^Y = 0$ satisfies the absence of a constant in Equation (93) as the first latent factor resembles a unit process under the risk-neutral measure for both variants. Imposing $\kappa_{22}^{Y,\mathbb{Q}} = \kappa_{33}^{Y,\mathbb{Q}}$ ensures the mean-reversion rates of the second and third factor to be equal under the $\mathbb{Q}$-measure. For the independent factor variant six more constraints arise from the implied diagonal structure on $\boldsymbol{K}^{\boldsymbol{Y},\mathbb{P}}$, whereas the last restriction defines the scaling of curvature on the slope under the risk-neutral measure as $\kappa_{23}^{Y,\mathbb{Q}} = \kappa_{22}^{Y,\mathbb{Q}}\left(\frac{\kappa_{13}^{Y,\mathbb{Q}}}{\kappa_{12}^{Y,\mathbb{Q}}}\right)$.

Discretization of the continuous-time parameters from Equation (23) allows for statistical inference via Gaussian state space methods (Christensen et al. (2011); Caldeira et al. (2016)). For the continuous-time AFRNS model, the mean and variance conditional on filtration $\mathcal{I}_t$ are

$$\mathbb{E}^{\mathbb{P}}[\boldsymbol{X}_T|\mathcal{I}_t] = \left(\boldsymbol{I} - \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}\right)\boldsymbol{\theta}^{\mathbb{P}} + \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}\boldsymbol{X_t}, \tag{24}$$

$$\mathbb{V}^{\mathbb{P}}[\boldsymbol{X_T}|\mathcal{I}_t] = \int_0^{\Delta t} \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}s}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\boldsymbol{e}^{-(\boldsymbol{K}^{\mathbb{P}})'s}\mathrm{d}s, \tag{25}$$

where $\Delta t = T - t$ in years. Consequentially, we exploit conditional moments of discrete observations to obtain both the measurement equation and state equation for the AFRNS model as

$$\boldsymbol{y_t} = \boldsymbol{a} + \boldsymbol{B}_{\boldsymbol{\lambda}}^*\boldsymbol{\beta_t} + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \tag{26}$$

$$\boldsymbol{\beta_t} = \left(\boldsymbol{I} - \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}\right)\boldsymbol{\theta}^{\mathbb{P}} + \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, \ldots, T, \tag{27}$$

where $\boldsymbol{a}$ $(N \times 1)$ refers to the deterministic yield-adjustment term with closed-form computation according to Appendix A.3, $\boldsymbol{B}_{\boldsymbol{\lambda}}^*$ $(N \times 3)$ resembles the adjusted base from Nyholm (2018), $\boldsymbol{\beta_t}$ (3×1) coincides with the latent states at time t to conform with the notation from Section 2.2, $\boldsymbol{H}$ $(N \times N)$ corresponds to a diagonal covariance matrix, $\boldsymbol{K}^{\mathbb{P}}$ and $\boldsymbol{\theta}^{\mathbb{P}}$ coincide with the continuous-time state dynamics and drift from Equation (23), respectively, and $\Delta t = 1/12$ refers to the annualized period between monthly observations.

Fisher & Gilles (1996) show that the imposition of no-arbitrage demands a special structure of $\boldsymbol{Q}$ (3×3), i.e.

$$\boldsymbol{Q} = \int_0^{\Delta t} \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}s}\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\boldsymbol{e}^{-(\boldsymbol{K}^{\mathbb{P}})'s}\mathrm{d}s, \tag{28}$$

which bridges the gap between the $\mathbb{P}$-measure and the $\mathbb{Q}$-measure as it is directly tied to the elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{K}^{\mathbb{P}}$. Moreover, we assume that the disturbances of the measurement equation and the state equation are orthogonal.

The state space representation of the AFRNS model is analogue to the multivariate formulations of earlier mentioned specifications. From one perspective, we obtain the RNS model via withdrawal of yield-adjustment $\boldsymbol{a}$ from Equation (26) while offering greater statistical freedom on the structure of the state dynamics $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$ and the state covariance matrix $\boldsymbol{Q}$. Alternatively, the state space representation of the AFNS model arises from two modifications on Equation (26), i.e. substitution of $\boldsymbol{B}_{\boldsymbol{\lambda}}^*$ with $\boldsymbol{B}_{\boldsymbol{\lambda}}$ and deterministic computation of AFNS's yield-adjustment instead of AFRNS's yield-adjustment (see Equation (88) in Appendix A.3).

2.4 Macro-Finance Rotated Nelson-Siegel

The facilitation of exogenous drivers often provides supplemental explanatory power to the term structure of interest rates. Prior to the rise of such literature, an unusually extensive gap between macroeconomists and financial economists remained apparent in their approaches to model the yield curve. The first group focuses on the role of inflation expectations and real macroeconomic activity, whereas the second group avoids the involvements of such determinants and resorts to the evolution of latent factors. Recent advances in literature unify both camps and demonstrate subordinate modelling improvements to emerge from the inclusion of observable components within the class of ATSMs (Ang & Piazzesi (2003); Kim & Wright (2005)), where after others show the surplus of explanatory power arising from the inclusion of exogenous factors to often persist with (modifications of) the DNS model (De Pooter et al. (2010); Yu & Zivot (2011); Mönch (2012)).

We bridge the gap between macroeconomist and financial economist yield curve modelling and depart from Section 2.2 to construct the Macro-Finance Rotated Nelson-Siegel (MF-RNS) model.[4] Following Diebold et al. (2006) we complement our latent state diffusion with P observable series in a nonstructural VAR(1) representation. Our proposed specification offers greater statistical freedom opposed to the exogenous connections from Nyholm (2018) as we allow for bi-directional dependencies between the latent states and the exogenous factors on the one hand and provide support for the exogenous factors to readily interfere with slope and curvature on the other hand. The state space composition of the MF-RNS model comprises

$$\begin{pmatrix} \boldsymbol{y_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \begin{bmatrix} \boldsymbol{B^*_\lambda} & \boldsymbol{0_{N\times P}} \\ \boldsymbol{0_{P\times 3}} & \boldsymbol{I_P} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon_t} \\ \boldsymbol{0_P} \end{pmatrix}, \quad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \tag{29}$$

$$\begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \left(\boldsymbol{I_{3+P}} - \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \right) \begin{pmatrix} \boldsymbol{\mu_\beta} \\ \boldsymbol{\mu_\gamma} \end{pmatrix} + \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_{t-1}} \\ \boldsymbol{\gamma_{t-1}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix}, \tag{30}$$

where $\boldsymbol{\gamma_t}$ $(P \times 1)$ bundles the exogenous factors at time t, $\boldsymbol{B^*_\lambda}$ $(N \times 3)$ refers to the factor loadings from Nyholm (2018) and covariance matrix $\boldsymbol{H}$ $(N \times N)$ remains diagonal. Matrices $\boldsymbol{K_1}$ (3×3), $\boldsymbol{K_2}$ $(3 \times P)$, $\boldsymbol{K_3}$ $(P \times 3)$ and $\boldsymbol{K_4}$ $(P \times P)$ describe the co-movements between the latent states and the exogenous variables. More specific, the blocks $\boldsymbol{K_1}$ and $\boldsymbol{K_4}$ show the within-set responses of the latent factors and the exogenous variables to their first lags, respectively, while the cross-sectional connections between the term structure of interest rates and the exogenous factors follow from $\boldsymbol{K_2} \neq \boldsymbol{0}$ and $\boldsymbol{K_3} \neq \boldsymbol{0}$. The first inequality portrays the yield curve's response to the exogenous factors (Kim & Wright (2005); Ang et al. (2006)), whereas the second inequality displays the term structure of interest rates as an indicator for future macroeconomic developments (Estrella & Mishkin (1996); Wright (2006)). The unconditional mean vector consists of two partitions, i.e. $\boldsymbol{\mu_\beta}$ (3×1) coincides with the unconditional means of the latent states and $\boldsymbol{\mu_\gamma}$ $(P \times 1)$ refers to the unconditional means of the exogenous factors. The assumed

[4] The nomenclature of MF-RNS stems from the lack of consensus between macroeconomists and financial economists to model the term structure of interest rates. Our approach unites both philosophies within the MF-RNS framework.

error structure for the state equation of the MF-RNS model is

$$\begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0}_3 \\ \mathbf{0}_P \end{pmatrix}, \begin{bmatrix} \boldsymbol{Q} & \mathbf{0_{3\times P}} \\ \mathbf{0_{P\times 3}} & \boldsymbol{R} \end{bmatrix} \right), \tag{31}$$

which restricts interrelated shocks on disturbances between the set of latent states and exogenous factors to be nonexistent. Similar to Sections 2.2 and 2.3 we define $\boldsymbol{Q}$ (3×3) as the covariance matrix for the latent factors. Additionally, we follow Diebold & Li (2006) and assume $\boldsymbol{R}$ $(P \times P)$ to be Cholesky-decomposable via a lower-triangular matrix. The analogy with an independent factor variant and a correlated factor variant of the MF-RNS model lies in the structure of $\boldsymbol{K_1}$ and $\boldsymbol{Q}$. The class of independent factors imposes a diagonal structure on $\boldsymbol{K_1}$ and $\boldsymbol{Q}$ while it leaves the structure of $\boldsymbol{K_2}$, $\boldsymbol{K_3}$, $\boldsymbol{K_4}$ and $\boldsymbol{R}$ untouched. Contrarily, a correlated factor variant of the MF-RNS model allows for a fully-flexible matrix $\boldsymbol{K_1}$ and a Cholesky-decomposable covariance matrix $\boldsymbol{Q}$.

The representation of the MF-RNS model in Equations (29) and (30) embodies a generalized formulation of earlier discussed specifications. We find the RNS model from Section 2.2 to be nested within our MF-RNS specification. Jointly restricting $\boldsymbol{K_2} = \mathbf{0}$ and $\boldsymbol{K_3} = \mathbf{0}$ partitions the state equation of the MF-RNS model in two sub-equations. Leaving out the VAR(1) process on the exogenous factors results in the RNS model. Additionally, the proposed exogenous connections of the theoretical short rate with inflation and capacity utilization from Nyholm (2018) fits our system via $P = 2$, a nonzero first row of $\boldsymbol{K_2}$ and $\boldsymbol{K_3} = \mathbf{0}$. On the other hand, one obtains the DNS-Macro model from Diebold et al. (2006) after substitution of the rotated factor loadings with the traditional Nelson-Siegel base from Equation (4).

2.5 Macro-Finance Arbitrage-Free Rotated Nelson-Siegel

We proposed two separate modifications to the baseline RNS model in Sections 2.3 and 2.4, respectively, although it appears beneficial to merge both in one specification. From one perspective, the MF-RNS model unites macroeconomist and financial economist yield curve modelling as it involves exogenous determinants, yet it lacks the theoretical rigour of the methods from Ang & Piazzesi (2003) and Kim & Wright (2005) as it fails to be arbitrage-free (Filipović (1999); Björk & Christensen (1999)). Alternatively, the AFRNS model from Section 2.3 complies with no-arbitrage and tractable estimation but ignores the possible surplus of explanatory power emerging from the modelled bi-directional dependencies between exogenous variables and the latent factors.

We combine both strands of literature in the Macro-Finance Arbitrage-Free Rotated Nelson-Siegel (MF-AFRNS) model, which closely follows the formulation of the MF-RNS specification from Section 2.4. Following Ullah (2016) we complement the discretized latent state diffusion of a no-arbitrage variant within the Nelson-Siegel environment with P exogenous components while our proposed MF-AFRNS specification preserves the factor-interpretation from Nyholm (2018). As Nyholm (2015) debates the entrance of observable series within the class of arbitrage-free Nelson-Siegel specifications to be unspanned, i.e. the incorporation of exogenous

factors should be limited to the state equation to not distort the arbitrage-free dynamics, the state space representation of the MF-AFRNS model settles to

$$\begin{pmatrix} \boldsymbol{y_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{0} \end{pmatrix} + \begin{bmatrix} \boldsymbol{B^*_\lambda} & \boldsymbol{0_{N\times P}} \\ \boldsymbol{0_{P\times 3}} & \boldsymbol{I_P} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon_t} \\ \boldsymbol{0_P} \end{pmatrix}, \quad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \tag{32}$$

$$\begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \left(\boldsymbol{I_{3+P}} - \begin{bmatrix} \boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \right) \begin{pmatrix} \boldsymbol{\theta_\beta} \\ \boldsymbol{\theta_\gamma} \end{pmatrix} + \begin{bmatrix} \boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_{t-1}} \\ \boldsymbol{\gamma_{t-1}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix}, \tag{33}$$

where the imposition of arbitrage-free dynamics introduces three modifications relative to the state space representation of the MF-RNS model. First, unification of the MF-RNS model with the no-arbitrage framework from Duffie & Kan (1996) requires the upper-left partition of the state dynamics, that is $\boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}}$ (3×3), to be of a specific structure in accordance with the discretization of the continuous-time state dynamics from Equation (23). Secondly, the imposed arbitrage-free conditions statistically limit the composition of $\boldsymbol{Q}$ (3×3) via

$$\boldsymbol{Q} = \int_0^{\Delta t} \boldsymbol{e^{-K_1^{\mathbb{P}}s}} \boldsymbol{\Sigma\Sigma'} \boldsymbol{e^{-(K_1^{\mathbb{P}})'s}} \mathrm{d}s, \tag{34}$$

although the assumed error distribution on the state equation of the MF-AFRNS model yields symbolic equivalence to Equation (31). The construction of $\boldsymbol{Q}$ via Equation (34) connects both measures as it involves the continuous-time state dynamics matrix $\boldsymbol{K_1^{\mathbb{P}}}$ and the risk-neutral volatility matrix $\boldsymbol{\Sigma}$. Thirdly, $\boldsymbol{a}$ $(N \times 1)$ coincides with the deterministic yield-adjustment term of the AFRNS model for which we provide an algebraic expression in Appendix A.3. All other parameters within the state space representation of the MF-AFRNS model yield equivalent interpretation and dimensionality to its counterparts from the MF-RNS model, although the different notation for the unconditional means arises from the no-arbitrage notation by Duffie & Kan (1996).

Our proposition of the MF-AFRNS model closely relates to earlier debated specifications. The AFNS-Macro specification from Ullah (2016), which combines the AFNS model with the inclusion of exogenous factors, emerges from substitution of the yield-adjustment and factor loadings with their AFNS-related counterparts (e.g. see Section 2.3). Alternatively, the MF-RNS specification arises from withdrawal of the yield-adjustment term while offering greater statistical flexibility on both $\boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}}$ and $\boldsymbol{Q}$. Lastly, we find the AFRNS model to be nested in Equations (32) and (33) as jointly restricting $\boldsymbol{K_2} = \boldsymbol{0}$ and $\boldsymbol{K_3} = \boldsymbol{0}$ divides the state equation of the MF-AFRNS model in two partitions. In line with the nested relation between the RNS and the MF-RNS specifications, simply ignoring the VAR(1) process on the exogenous factors yields the AFRNS model.

3 Estimation

We estimate our discussed specifications using Bayesian Markov Chain Monte Carlo (MCMC) techniques as they overcome several caveats arising from frequentist methods. Our estimation

approach circumvents the problematic convergence to local maxima, which traditionally accompanies classical no-arbitrage methods (Kim & Wright (2005)). Also, it evades the onerous and costly initialization routines for arbitrage-free specifications with the traditional Nelson-Siegel base and exogenous factors from Ullah (2016). Furthermore, our approach methodologically surpluses frequentist estimation methods as it incorporates parameter- and stochastic uncertainty (De Pooter et al. (2007)). Lastly, our approach is exact and does not require asymptotic distribution theory to assess the properties of the estimators.

3.1 In-Sample Posterior Simulation

We generalize notation to fit our discussed specifications from Section 2 and express a standard linear Gaussian state space model as

$$\boldsymbol{y_t} = \boldsymbol{d} + \boldsymbol{Z}\boldsymbol{\alpha_t} + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \tag{35}$$

$$\boldsymbol{\alpha_{t+1}} = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{\alpha_t} + \boldsymbol{\eta_t}, \qquad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, \dots, T, \tag{36}$$

where $\boldsymbol{y_t}$ $(N \times 1)$ refers to the observed yields at time t, $\boldsymbol{\alpha_t}$ $(J \times 1)$ coincides with the state vector at time t and $\boldsymbol{d}$ $(N \times 1)$, $\boldsymbol{Z}$ $(N \times J)$, $\boldsymbol{H}$ $(N \times N)$, $\boldsymbol{c}$ $(J \times 1)$, $\boldsymbol{T}$ $(J \times J)$, $\boldsymbol{Q}$ $(J \times J)$ correspond to the parameters of the model respectively. The state space representation from Equations (35) and (36) resolves to a time-invariant version of the model proposed by Durbin & Koopman (2012) as we suppress the possible time-dependence in the parameters.

Our Bayesian MCMC algorithm exploits Gibbs sampling and Data Augmentation to first account for stochastic uncertainty in the latent factors conditional on the sampled parameters. Hereafter, we reverse the procedure to sample the parameters from their full conditional marginal posterior distributions given the latent states (Geman & Geman (1984); Tanner & Wong (1987); Carter & Kohn (1994)).[5] We define $\Theta = \{\boldsymbol{d}, \boldsymbol{Z}, \boldsymbol{H}, \boldsymbol{c}, \boldsymbol{T}, \boldsymbol{Q}\}$ as the set of parameters, $\boldsymbol{\alpha_{1:T}}$ as the latent states and $\boldsymbol{y_{1:T}}$ as all in-sample observations. Then, Bayes' rule implies the joint posterior of Θ and $\boldsymbol{\alpha_{1:T}}$ to be of the form

$$\begin{aligned} p(\Theta, \boldsymbol{\alpha_{1:T}} | \boldsymbol{y_{1:T}}) &\propto p(\Theta) p(\boldsymbol{y_{1:T}}, \boldsymbol{\alpha_{1:T}} | \Theta) \\ &\propto p(\Theta) p(\boldsymbol{y_{1:T}} | \boldsymbol{\alpha_{1:T}}, \Theta) p(\boldsymbol{\alpha_{1:T}} | \Theta), \end{aligned} \tag{37}$$

where $p(\Theta)$ refers to a product of independent prior densities, $p(\boldsymbol{y_{1:T}}, \boldsymbol{\alpha_{1:T}} | \Theta)$ coincides with the complete data likelihood and $p(\boldsymbol{\alpha_{1:T}} | \Theta)$ allows the researcher to impose prior knowledge on the states. With Data Augmentation posterior inference simplifies to standard regression results for both the measurement equation and the state equation as it requires us to evaluate $p(\boldsymbol{y_{1:T}} | \boldsymbol{\alpha_{1:T}}, \Theta)$, i.e. the likelihood conditional on the latent factors. If we define $\Theta^{(m)}$ and $\boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}$ as the m-th draw of the model's parameters and states respectively, the Gibbs sampler of joint posterior $p(\Theta, \boldsymbol{\alpha_{1:T}} | \boldsymbol{y_{1:T}})$ takes the following shape:

[5] We follow Hautsch & Yang (2012) and Çakmaklı (2020) in their approaches and explicitly omit the first observation in the construction of the posterior kernel for the state equations. This omission yields greater simplicity as it preserves the validity of the Gibbs sampler over the Metropolis-Hastings algorithm (see Hastings (1970)).

- Initialize parameters $\Theta^{(0)}$ and states $\boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(0)}$. Then, at the m-th iteration:
  1. Sample $\boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}$ from $p(\boldsymbol{\alpha}_{\boldsymbol{1:T}}|\Theta^{(m-1)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
  2. Sample $\boldsymbol{H}^{(m)}$ from $p(\boldsymbol{H}|\boldsymbol{Z}^{(m-1)}, \boldsymbol{d}^{(m-1)}, \boldsymbol{Q}^{(m-1)}, \boldsymbol{T}^{(m-1)}, \boldsymbol{c}^{(m-1)}, \boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
  3. Sample $\boldsymbol{Z}^{(m)}$ from $p(\boldsymbol{Z}|\boldsymbol{H}^{(m)}, \boldsymbol{d}^{(m-1)}, \boldsymbol{Q}^{(m-1)}, \boldsymbol{T}^{(m-1)}, \boldsymbol{c}^{(m-1)}, \boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
  4. Sample $\boldsymbol{d}^{(m)}$ from $p(\boldsymbol{d}|\boldsymbol{H}^{(m)}, \boldsymbol{Z}^{(m)}, \boldsymbol{Q}^{(m-1)}, \boldsymbol{T}^{(m-1)}, \boldsymbol{c}^{(m-1)}, \boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
  5. Sample $\boldsymbol{Q}^{(m)}$ from $p(\boldsymbol{Q}|\boldsymbol{H}^{(m)}, \boldsymbol{Z}^{(m)}, \boldsymbol{d}^{(m)}, \boldsymbol{T}^{(m-1)}, \boldsymbol{c}^{(m-1)}, \boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
  6. Sample $\boldsymbol{T}^{(m)}$ from $p(\boldsymbol{T}|\boldsymbol{H}^{(m)}, \boldsymbol{Z}^{(m)}, \boldsymbol{d}^{(m)}, \boldsymbol{Q}^{(m)}, \boldsymbol{c}^{(m-1)}, \boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
  7. Sample $\boldsymbol{c}^{(m)}$ from $p(\boldsymbol{c}|\boldsymbol{H}^{(m)}, \boldsymbol{Z}^{(m)}, \boldsymbol{d}^{(m)}, \boldsymbol{Q}^{(m)}, \boldsymbol{T}^{(m)}, \boldsymbol{\alpha}_{\boldsymbol{1:T}}^{(m)}, \boldsymbol{y}_{\boldsymbol{1:T}})$.
- Repeat (1)-(7) for a user-specified amount of M times.

We resort to the simulation smoother from Durbin & Koopman (2002) for posterior inference on the latent states in the first step of our algorithm. Their multivariate method is highly efficient as it suffices to run both the Kalman Filter and the Kalman Smoother once per iteration (e.g. see Appendix C for an explanation in greater detail). Meanwhile, it also results in faster convergence of the MCMC chains as it reduces the autocorrelations of the sampled states in comparison to the element-wise recursive techniques from Carter & Kohn (1994) and Frühwirth-Schnatter (1994). For steps (2)-(6) of our algorithm, we resort to multivariate sampling distributions as Chib & Greenberg (1995) find joint sampling of strongly correlated parameters to account for rapid movement towards regions of high posterior probability. The sampling algorithm complies with the order from Hautsch & Yang (2012) as we chronologically sample the states, the measurement equation and the state equation. Per equation, we first sample the covariances after which we sample the state dynamics.

In Appendices D and F, we admit both the RNS and the MF-RNS specifications to the general notation from Equations (35) and (36) while we also provide the full conditional marginal posteriors for all parameters considered. The sampling procedures for both specifications fit the Gibbs sampler via the elimination of two steps, i.e. we exclude steps (3) and (4) as the deterministic matrix $\boldsymbol{B}_{\boldsymbol{\lambda}}^{*}$ takes the place of $\boldsymbol{Z}$ in Equation (35) and both models feature no constant in their measurement equations.[6]

The imposition of no-arbitrage results in a deviating fit for both the AFRNS and the MF-AFRNS model to the Gibbs sampler. In line with the mappings of our sole statistical specifications, we exclude step (3) from the respective sampling schemes of the AFRNS and the MF-AFRNS model as the deterministic completion of $\boldsymbol{B}_{\boldsymbol{\lambda}}^{*}$ extends to our arbitrage-free specifications. The difference lies in the handling of step (4), i.e. both the RNS and the MF-RNS model exclude this step from their sampling schemes as they have no constant in their respective measurement equations. The imposition of no-arbitrage does add a constant to the measurement equation in form of the yield-adjustment. However, randomly drawing the yield-adjustment from the posterior density of step (4) is invalid as its completion is deterministic conditional on the risk-neutral volatility matrix $\boldsymbol{\Sigma}$ and the rate of decay λ (e.g. see Appendix A.3). Hence, we

[6]The deterministic property of $\boldsymbol{B}_{\boldsymbol{\lambda}}^{*}$ follows from the fact that we plug-in λ via the two-step estimation method from Diebold & Li (2006) and hold it fixed throughout the entire sampling routine (e.g. see Section 3.3).

exploit the untethered $\mathbb{P}$-dynamics of both the AFRNS and the MF-AFRNS model to ex post compute the yield-adjustment as an additional step to the Gibbs sampler. More specific, we omit step (4) while we inversely compute the continuous-time state dynamics and the risk-neutral volatility matrix with the freely-sampled discretized state dynamics and covariance matrix from steps (5) and (6) respectively.

We illustrate our Bayesian approach with the incorporation of no-arbitrage dynamics in the AFRNS model, although we refer to Appendix E for an explanation in greater detail. Following the steps of the Gibbs sampler, we randomly draw samples of $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$ and $\boldsymbol{Q}$ in steps (5) and (6). Next, we extract $\boldsymbol{K}^{\mathbb{P}}$ via element-wise inverse operations on $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$. Hereafter, we compute $\boldsymbol{\Sigma}$ via the inverse formulation of Equation (28). We subsequently insert the computed $\boldsymbol{\Sigma}$ in Equation (75) to compute the yield-adjustment term as a final step of the Gibbs sampler, which reversely ties the no-arbitrage property to the AFRNS model.[7]

Our Bayesian completion on the imposition of no-arbitrage within the Nelson-Siegel environment corresponds to an opposite approach of frequentist methods. Classical estimation suggests treating $\boldsymbol{K}^{\mathbb{P}}$ and $\boldsymbol{\Sigma}$ as arguments to enter the Nelder-Mead derivative-free optimization routine (e.g. see Singer & Nelder (2009) for an explanation in greater detail). Consequentially, the elements of $\boldsymbol{\Sigma}$ enter the yield-adjustment via Equation (75), $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$ follows numerically from the scaling and squaring method with Padé approximations by Al-Mohy & Higham (2010) and $\boldsymbol{Q}$ follows from Equation (28). In Appendix C.3, we elaborate on the estimation methods from Christensen et al. (2011) and Ullah (2016) with which we additionally estimate our specifications to evaluate the adequacy of our proposed Bayesian techniques.

Our Bayesian manner to unify the (MF-)RNS model with no-arbitrage dynamics has its limitations. Although the methods from Al-Mohy & Higham (2010) can be successfully implemented to extract a non-diagonal matrix $\boldsymbol{K}^{\mathbb{P}}$ from $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$, Fisher & Gilles (1996) show no exact solution to exist for subtracting a lower-diagonal matrix $\boldsymbol{\Sigma}$ from a Cholesky-decomposable matrix $\boldsymbol{Q}$. More specific, a hypothetical inverse formulation of Equation (28) requires $\boldsymbol{\Sigma}$ and $\boldsymbol{e}^{-(\boldsymbol{K}^{\mathbb{P}})s}$ to commute $\forall s$. As this assumption does not hold for correlated factor variants of the (MF-)AFRNS model, we leave the latter's implementation for further research.

3.2 Forecast Composition

Opposed to classical methods, Bayesian inference sustains the incorporation of parameter-, forecast-, and stochastic uncertainty via predictive densities rather than to rely on point forecasts solely. Given the generalized Gaussian state space model from Equations (35) and (36), a closed-form expression for one-step-ahead predictive density $p(\boldsymbol{y_{T+1}}|\boldsymbol{y_{1:T}})$ yields

$$\begin{aligned} p(\boldsymbol{y_{T+1}}|\boldsymbol{y_{1:T}}) &= \int\int p(\boldsymbol{y_{T+1}}, \Theta, \boldsymbol{\alpha_{1:T}}|\boldsymbol{y_{1:T}})\mathrm{d}\Theta\mathrm{d}\boldsymbol{\alpha_{1:T}} \\ &= \int\int p(\boldsymbol{y_{T+1}}|\Theta, \boldsymbol{\alpha_{1:T}}, \boldsymbol{y_{1:T}})p(\Theta, \boldsymbol{\alpha_{1:T}}|\boldsymbol{y_{1:T}})\mathrm{d}\Theta\mathrm{d}\boldsymbol{\alpha_{1:T}}, \end{aligned} \tag{38}$$

[7] The same principle applies to the MF-AFRNS model, i.e. we extract $\boldsymbol{K}_{\boldsymbol{1}}^{\mathbb{P}}$ from $\boldsymbol{e}^{-\boldsymbol{K}_{\boldsymbol{1}}^{\mathbb{P}}\Delta t}$ while we compute $\boldsymbol{\Sigma}$ via the inverse formulation of Equation (34) with a freely-sampled covariance matrix $\boldsymbol{Q}$ (e.g. see Appendix G for an explanation in greater detail).

where $p(\boldsymbol{y_{T+1}}|\Theta, \boldsymbol{\alpha_{1:T}}, \boldsymbol{y_{1:T}})$ resembles the one-step-ahead forecast distribution conditional on the parameters and latent factors and $p(\Theta, \boldsymbol{\alpha_{1:T}}|\boldsymbol{y_{1:T}})$ refers to the in-sample joint posterior density from Section 3.1. Unfortunately, the obtained algebraic expression for the one-step-ahead forecast density from Equation (38) has limited practical scope as it relies on diffuse or natural conjugate priors and there is no such expression for larger forecast horizons.

Hence, we resort to the simulation techniques from Litterman (1986) as he offers wider applicability while accounting for parameter- and stochastic uncertainty. Following his methods, we construct predictive density $p(\boldsymbol{y_{T+h}}|\boldsymbol{y_{1:T}})$ via additional steps to our algorithm, i.e.

- Obtain parameters $\Theta^{(m)}$ and states $\boldsymbol{\alpha}^{(m)}_{\boldsymbol{1:T}}$ from steps (1)-(7). Then, for $i = 1, 2, \dots h$:
  8. Compute $\boldsymbol{\alpha}^{(m)}_{\boldsymbol{T+i}} = \boldsymbol{c}^{(m)} + \boldsymbol{T}^{(m)}\boldsymbol{\alpha}^{(m)}_{\boldsymbol{T+i-1}} + \boldsymbol{\eta}^{(m)}_{\boldsymbol{T+i}}$, where $\boldsymbol{\eta}^{(m)}_{\boldsymbol{T+i}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{(m)})$.
  9. Compute $\boldsymbol{y}^{(m)}_{\boldsymbol{T+i}} = \boldsymbol{Z}^{(m)}\boldsymbol{\alpha}^{(m)}_{\boldsymbol{T+i}} + \boldsymbol{\varepsilon}^{(m)}_{\boldsymbol{T+i}}$, where $\boldsymbol{\varepsilon}^{(m)}_{\boldsymbol{T+i}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}^{(m)})$.
- Repeat (1)-(9) for a user-specified amount of M times.

We refer to the m-th draw of predictive density $p(\boldsymbol{y_{T+i}}|\boldsymbol{y_{1:T}})$ with $\boldsymbol{y}^{(m)}_{\boldsymbol{T+i}}$. We involve parameter uncertainty in the construction of our forecasts as the point estimates to be reported refer to the posterior means of the respective predictive densities. Additionally, we account for stochastic uncertainty as we randomly draw $\boldsymbol{\eta}^{(m)}_{\boldsymbol{T+i}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{(m)})$ and $\boldsymbol{\varepsilon}^{(m)}_{\boldsymbol{T+i}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}^{(m)})$ in steps (8) and (9) for all iterations and forecasts considered.

3.3 Initialization Routine & Prior Distributions

We exclude λ from the set of parameters to preliminary compute it via

$$\begin{aligned} &\Leftrightarrow \quad \text{argmin}_{\lambda, \{\boldsymbol{\beta}_t\}_{t=1}^T} \sum_{t=1}^{T} \Big(\boldsymbol{y_t} - \boldsymbol{B^*_\lambda}\boldsymbol{\beta_t}\Big)' \Big(\boldsymbol{y_t} - \boldsymbol{B^*_\lambda}\boldsymbol{\beta_t}\Big), \\ &\Leftrightarrow \quad \text{argmin}_{\lambda} \sum_{t=1}^{T} \Big(\boldsymbol{y_t'}\boldsymbol{y_t} - \boldsymbol{y_t'}\boldsymbol{B^*_\lambda}\Big(\boldsymbol{B^{*\prime}_\lambda}\boldsymbol{B^*_\lambda}\Big)^{-1}\boldsymbol{B^{*\prime}_\lambda}\boldsymbol{y_t}\Big), \end{aligned} \tag{39}$$

where the optimization reduction follows from orthogonal projectors. A deterministic plug-in of λ simplifies identification while it provides greater robustness for the posterior results (Çakmaklı (2020)). The initialization of our routine follows from the two-step estimation method by Diebold & Li (2006), i.e. we construct $\Theta^{(0)}$ via the OLS estimates of the measurement equation and the state equation, respectively. Additionally, we ensure the state equations for all specifications considered to be covariance stationary, i.e. we constrain the real moduli of the state dynamics' eigenvalues to be larger than zero for our arbitrage-free specifications while we constrain the eigenvalues of the state dynamics matrices for sole statistical specifications to be below one. For arbitrage-free configurations, we forge augmented starting values via alignment of the two-step estimates with the specified structure of $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$ and $\boldsymbol{Q}$. [8]

Although inconsistent with the true spirit of Bayesian analysis we follow De Pooter et al. (2007), Diebold et al. (2008) and Hautsch & Ou (2008) to design prior densities with knowledge

[8] We use the same procedures as within our Bayesian sampling routine to extract $\boldsymbol{K}^{\mathbb{P}}$ and $\boldsymbol{\Sigma}$ from $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}$ and $\boldsymbol{Q}$ respectively

on the empirical behavior of the latent states. Although all other priors are set diffuse, we follow Diebold et al. (2008) and impose a $\mathcal{N}(\bar{\boldsymbol{\beta}}, 0.01)$ prior on the unconditional means, where $\bar{\boldsymbol{\beta}}$ refers to the sample means of the latent factors from the initial two-step approach. We found the imposition of these data-driven priors on the unconditional means to greatly facilitate estimation. Absence of these priors have been found to result in unusual behavior of the sampled unconditional means as large posterior variances emerged from high degrees of persistence. The true nature of Bayesian analysis would be preserved if we calibrate the prior parameters via pre-sample data. However, a similar setting appears infeasible to our empirical exercise as there is no pre-sample data available for our supply factor (e.g. see Section 4). Besides, the advantages of data-driven priors are multi-fold. The imposition of these supports a replicate manner to strengthen the acquired estimates, whereas alternative prior selection may resolve to mere guesswork. Moreover, one could argue if the implementation of prior research is indeed in the true spirit of Bayesian analysis if the examined sample periods overlap.

4 Data

In Section 4.1, we elaborate on the merged construction of our empirical yield data set, which combines observed components from two databases. We find the prolonged period of near-zero rates to have alternated many of the stylized facts on the yield curve. Nevertheless, a data-driven exercise provides empirical support for our discussed specifications to adequately capture the changed dynamics. Our constructed set of observed variables finds inspiration in the suggested exogenous regressors from Diebold et al. (2006). They use the Federal Funds Rate (FFR) as a monetary policy variable in addition to approximates for inflation and economic activity. We replace the FFR with the proposed monetary policy instrument from Section 4.2 because the FFR is limited to the Zero Lower Bound (ZLB) in recent times and monetary institutions seek to accommodate further stimulus via the constructed supply factor from Section 4.2. For the other two variables, we follow Diebold et al. (2006) and complement the set of exogenous factors with empirical approximates for inflation and economic activity, respectively, in Section 4.3.

4.1 Yields

If the U.S. Department of Treasury were to supply a continuous spectrum of zero-coupon bonds, the yield curve would be observed. However, the issued amount of distinct maturities is limited and coupon bonds arise. Even though the latter follows the interpretation of a basket with modest zero-coupon securities and a final principal payment, methods as bootstrapping do not suffice as the observed discrete securities do not span the entire maturity spectrum.

Therefore, we obtain end-of-month zero-coupon yields from July 2003 to February 2020 with maturities from three months up to ten years, i.e. $\tau_i \in \{3, 6, 12, 24, 36, 60, 84, 120\}$ in months.[9] The Federal Reserve publishes both observable yields and U.S. Treasury OLS estimates for the Svensson curve daily in their Gürkaynak et al. (2007) database, where we extract the visible

[9]The short in-sample period stems from limited availability of our supply factor, i.e. the first releases of both the SOMA and MSPD reports were in July 2003 (e.g. see Section 4.2 for an explanation in greater detail).

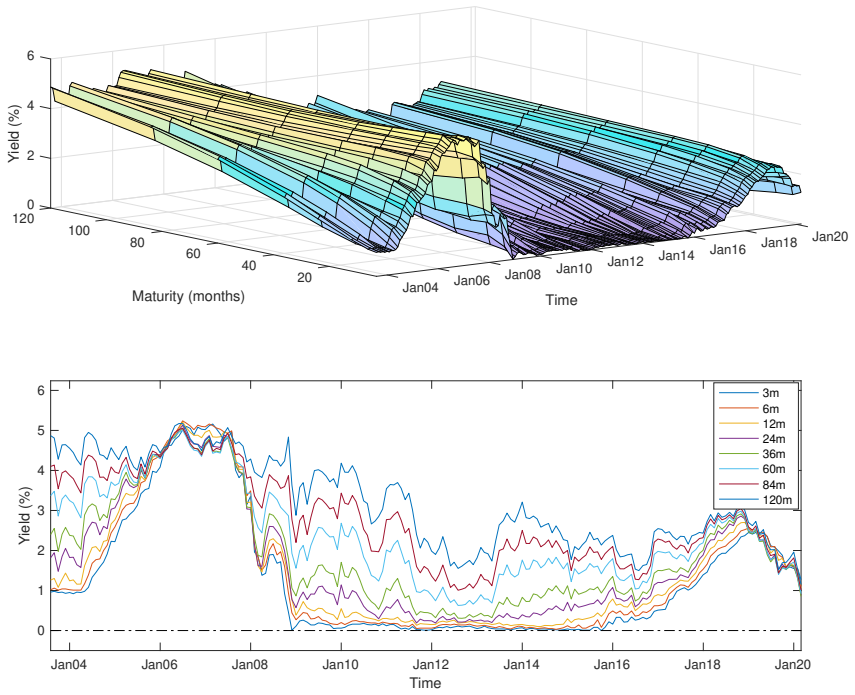

Figure 3: Time series of U.S. yields over the period July 2003 up until February 2020 for the merged empirical data set with maturities $\tau_i \in \{3, 6, 12, 24, 36, 60, 84, 120\}$ in months. Maturities greater or equal to one year stem from the Gürkaynak et al. (2007) database, whereas we extract the shorter-term yields from the H.15 series.

components with maturities greater than six months. For shorter-term maturities, we follow Christensen & Rudebusch (2016) and resort to the H.15 series of the Federal Reserve board.[10]

Panel A of Table 2 shows many stylized facts to hold for our empirical exercise, although the ZLB period changed some dynamics for the entire cross-section of observed yields. We find an upward sloping concave shape on average and a near-decreasing pattern of standard deviations with maturity, i.e. the largest (smallest) standard deviations locate in the short-end (long-end) of the term structure. Contrarily, three features of our observed set of yields indicate the unusual interest rate environment. First, we observe a decreasing pattern of first-order correlations as maturity increases, whereas an increasing evolution is to be regularly expected. The unusual pattern arises from the persistent short-end of the term structure and finds graphical support in Figure 3. Secondly, we find the short-end of the term structure to approach the ZLB from November 2008 to October 2015 in response to unconventional monetary policy-making, although the positive minima in Table 2 assure our data set does not conflict with the non-zero truncation Black (1995). Thirdly, we find the upward sloping concave shape of the mean yields to not extend to the observed maxima as Figure 3 shows the historically-high flat yield curve before the Great Recession to be its cause.

Panel B of Table 2 elaborates on the summary statistics for the empirical approximates of the RNS model's latent factors. An observed estimate for the first latent factor stems from the rate

[10]The nominal NSS yields from Gürkaynak et al. (2007) stem from here, whereas the H.15 series may be assessed here. Missing observations were replaced with nearest available end-of-month yields.

Table 2: Descriptive statistics on the yield curves

Maturity	Mean	St. dev.	Min.	Max.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
Panel A: Observed yields							
3	1.318	1.570	0.000	5.160	0.993	0.738	0.267
6	1.435	1.598	0.030	5.240	0.994	0.754	0.306
12	1.547	1.534	0.099	5.210	0.993	0.769	0.354
24	1.725	1.438	0.188	5.130	0.989	0.789	0.433
36	1.946	1.333	0.306	5.058	0.985	0.800	0.487
60	2.391	1.178	0.627	5.008	0.979	0.802	0.557
84	2.775	1.110	1.007	5.052	0.975	0.796	0.611
120	3.205	1.096	1.149	5.172	0.973	0.799	0.681
Panel B: Empirical approximates							
Short rate	1.318	1.570	0.000	5.160	0.993	0.738	0.267
Slope	1.886	1.244	-0.550	4.375	0.973	0.554	-0.116
Curvature	-1.074	0.852	-2.691	0.366	0.957	0.632	0.279

Note: This table contains the means, standard deviations, minima, maxima and several periodical autocorrelations for each maturity and empirical approximate considered over the period July 2003 up until February 2020. All statistics are computed with yields in percentages and maturities in months. The periodical autocorrelations are given in months, i.e. $\hat{\rho}(1)$ refers to the monthly autocorrelation, $\hat{\rho}(12)$ corresponds to the annual autocorrelation and $\hat{\rho}(24)$ coincides with the two-year autocorrelation. The empirical approximates follow the construction provided in the main text.

with the lowest maturity, i.e. the three-month yield in our empirical data set. The conjecture on the slope follows the opposed sign relation from Frankel & Lown (1994) (e.g. see Section 2.2), where we subtract the empirically observed short rate from the ten-year yield. Lastly, we construct an approximate curvature factor via $2 \times y(24) - y(3) - y(10)$ where $y(i)$ denotes the time series of the yields with i months to maturity (Diebold & Li (2006)).

Although the current environment relegated some of the stylized facts on the term structure, we find empirical evidence for a three-factor state diffusion to sufficiently capture the underlying yield dynamics. Litterman (1991) and Wright (2011) mention that a three-factor latent state diffusion adequately captures time-variation in the cross-section of U.S. treasures. Indeed, we find that the first three principal components account for 99.9% of the total variation in term structure movements (see Figure 19 in Appendix H.2 for an explanation in greater detail). As these follow the interpretation of DNS's latent factors and RNS yields observational equivalence to DNS, we postulate (variations of) RNS to be well-suited for our empirical exercise.

4.2 Supply Factor

In the aftermath of the Great Recession, it initially posed a challenge to further combat the deteriorating economic outlook. The Federal Funds Rate (FFR) cut slack as the target for the overnight rate had been set to a range of 0 to 25 basis points by the Federal Open Market Committee (FOMC) in November 2008. Consequentially, the Federal Reserve started acquiring vast quantities of bonds with medium-term and long-term maturities via Large Scale Asset

Purchases (LSAPs). The programs intended to additionally suppress private borrowing costs and to revitalize the economy via improved credit market conditions. Three weeks later, the FOMC announced purchases of mortgage-backed securities (MBS) and housing agency debt of up to $600 billion. By March 2009, the FOMC declared a broad expansion of their purchases in agency-related securities and long-term treasuries as it held $1.75 trillion in notes, MBS and bank debt. In response to anemic economic growth and naturally falling holdings as debt matured, the FOMC embarked on the second round of Quantitative Easing (QE) in November 2010 and announced a further $600 billion expansion of long-term treasuries by Fall 2011. The third round of purchases was announced in September 2012 to increasingly restrain mortgage rates and to jolt the housing market via a monthly purchase program of $40 billion in MBS. In October 2017, the Federal Reserve Board declared a balance sheet normalization program to significantly lower the magnitude of reserve balances over the next few years (Bonis et al. (2017)).

The literature considers related expansionary monetary policy controversial and questions its effectiveness via a revitalized interest in a previously occurred comparable phenomenon. Operation Twist, a 1961 initiative undertaken by the Kennedy Administration and the Federal Reserve Board, aimed to clock-wise rotate the yield curve via enhancement of the short-end and suppression of the long-end. Ross (1966), Modigliani & Sutch (1966) and Holland (1969) consider the effectiveness of Operation Twist moderate at best, where Solow et al. (1987) and Blinder (2000) mention the insufficient size of the purchase programs to be the root cause of its lack of success. Similar findings are in line with no-arbitrage literature, which implies nonexisting supply effects as treasuries can be replicated using similar assets. Modern re-evaluation of Operation Twist yields significant but mild effects on the term structure of interest rates. Swanson et al. (2011) exploit intra-day data to find a statistically substantial but mediocre fifteen basis points reduction on the long-end of the term structure. They blame the found statistical insignificance of early research to have suffered from inherent lower-frequency problems.

Various researches show quantitative assessments of emerged supply effects in modern-day successors of Operation Twist, even though the current LSAPs limit their focus on the long-end of the yield curve. Laubach (2009) already discovered significant effects of government deficits and debt on interest rates before the Great Recession. Krishnamurty & Vissing-Jorgensen (2011) and Ihrig et al. (2012) find statistical evidence of significant downward pressure on long-term rates by LSAPs. Alternatively, Li & Wei (2012) explore the effects of supply factors on the term structure in the framework from Ang & Piazzesi (2003) to find a one percentage point reduction on ten-year treasury yields from the first two LSAPs. These findings suggest the incorporation of supply levels to possibly offer greater comprehension of future term structure advances.

Theorems on asset-specific supply and demand mechanisms provide a rationale for these significant findings as they contradict existing no-arbitrage literature. Specifically, the preferred habitat view from d'Amico et al. (2012) and the scarcity channel from Modigliani & Sutch (1966) motivate the use of LSAPs as a monetary policy instrument. The preferred habit view states that investors possess maturity in-elastic preferences due to their risk aversion, i.e. investors wish to be compensated for investing in deviating horizons as it exposes them to interest rate risks.

Per illustration, pension funds prefer to hold long-term bonds for their long-term liabilities and money market mutual funds prefer to hold short-term bonds to provide liquidity within their portfolio. The scarcity channel, or equivalently portfolio balance effect, states bonds to be isolated per maturity. A change in the supply of bonds with a particular maturity only affects yields of bonds with similar characteristics. As there exists no perfect substitution for yields with dissimilar maturities, a respective shock in supply or demand can not be arbitraged away.

Additionally, the alternated market expectations emerging from unconventional policy-making provide further theoretical support for QE. Markets not only respond to actual supply or demand shifts as bond prices can also fluctuate via changed expectations on policymakers' future proceedings. Specifically, recent LSAP announcements suppress the entire cross-section of yields via the signalling channel or the forward guiding principle, i.e. regulatory statements of extended LSAPs exert further downward pressure on long-term rates via an expanded period of future near-zero short rates (Bauer & Rudebusch (2013)). Incorporation of these alternated market expectations gives rise to the involvement of more lags. Nevertheless, we insist on the parsimony from the (V)AR(1) structure in traditional DNS modelling.

Following these empirical and theoretical underpinnings, we construct the Marketable Fraction (MF) of long-term treasuries as the monetary policy instrument within our set of variables.[11] Its construction follows from monthly releases of the Treasury's Monthly State of the Public Debt (MSPD) and the System of Open Market Account (SOMA). The former announces the total amount of outstanding treasuries, whereas the latter reports the amount of public debt in the Fed's portfolio. We filter both reports on CUSIPs with maturities greater than 9.5 years to compute the absolute amount of long-term public debt available via the MSPD while we compute the absolute amount of long-term public debt in the Fed's portfolio via the SOMA.[12] Then, a time series on the MF follows from one minus the ratio of the absolute amount of long-term securities on the Fed's balance sheet to the absolute amount of long-term treasuries publicly available.[13]

Figure 4 shows the Fed's response to accommodate further economic stimuli as the FFR reached its ZLB. Between the announcement dates of LSAP1 and LSAP2 the MF appears stable as both the long-term holdings in the SOMA and the long-term holdings in the MSPD approximately show the same relative growth. Beyond the announcement date of LSAP2, we find the relative growth of the long-term assets in the SOMA reports to outpace the growth of the publicly available debt in the MSPD reports, which lead to an approximate fifteen percentage point reduction of the MF by the announcement date of LSAP3. The MF ultimately decreases to 60.4% by August 2014, whereafter it rises due to stable long-term holdings in the SOMA reports and rising long-term holdings in the MSPD reports respectively.

[11] We limit our focus on the relative holdings of bonds with maturities greater than 9.5 years as LSAPs were originated to exert downward pressure on the long-end of the curve.

[12] A web-scraping tool to filter both reports on long-term securities was readily available at Rabobank in R, although it required to be updated. Hence, we would like to thank Rabobank for the permission to use their tool in this paper.

[13] The series appears to be stationary as we reject the null-hypothesis of the Augmented Dickey-Füller (ADF) test at a 95% confidence level with no lags included.

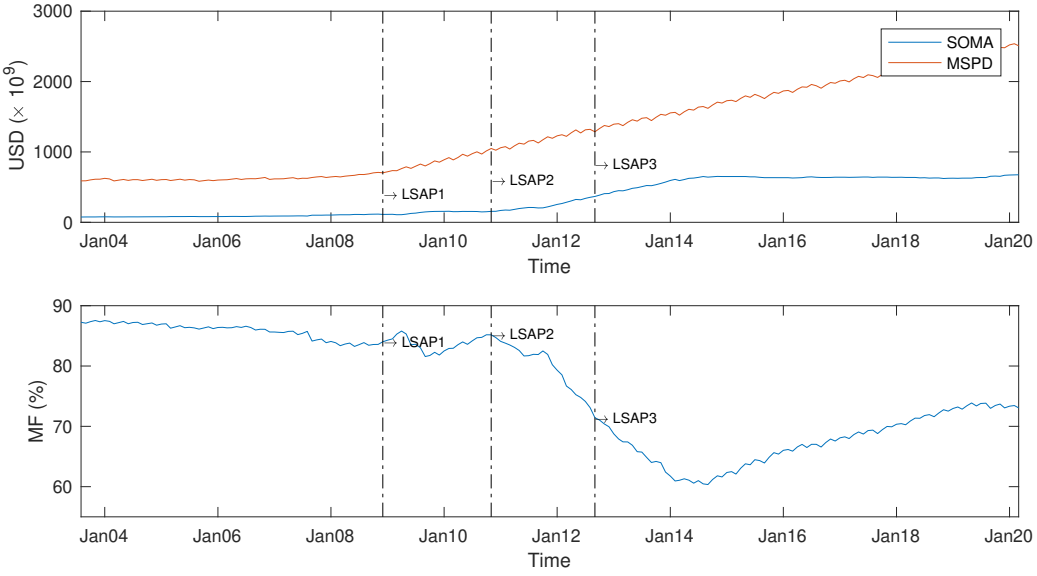

Figure 4: This figure consists of two partitions. The first partition shows the time series of long-term securities in the System of Open Market Account (SOMA) and the time series of long-term securities in the Monthly State of the Public Debt (MSPD). The second partition shows the time series of our monetary policy variable, the Marketable Fraction (MF) of long-term securities, which follows the construction mentioned in the main text. The vertical lines with subscript LSAPi, $i \in \{1, 2, 3\}$, refer to the respective announcement dates of the Large Scale Asset Purchases (LSAPs). All time series are obtained over the period July 2003 - February 2020.

4.3 Macro Factors

We complement our set of variables with approximates for economic growth and inflation, as we wish to include traditional dependencies of these macroeconomic variables with the short-end of the term structure in our models (Nyholm (2018)).

We consider annual logarithmic changes of inflation via the Core Consumer Price Index (CCPI), which circumvents instability via the exclusion of volatile price components (Clark et al. (2001)) and theoretical nonstationarity via its transformation (Charemza et al. (2005)). As an approximate for economic activity, we follow Bech & Lengwiler (2012) and seek resort to annual logarithmic growth of NonFarm Payroll (NFP). The obvious candidate would be GDP growth. However, the practicality of measures for GDP is limited due to quarterly publications that usually come one month after their respective reference quarters. NFP, on the other hand, is published every month with a public delay of four days. Moreover, Altavilla, Giannone, & Modugno (2017) show that 98% of evaluated Treasury bond market participants set an alarm for a newly scheduled release of the NFP figure in Bloomberg, which portrays its use among practitioners as an approximate for economic activity. Thus, we obtain the series on the CCPI and the NFP from the Federal Reserve Economic Data (FRED) database.[14]

Figure 5 shows plummeting NFP and deviating CCPI from the 2% trend with the fall of Lehman Brothers in 2008. Subsequently, policymakers sought to accommodate economic growth and to achieve their inflation targets via depression of the overnight rate. Originally, the monetary mechanism between these macroeconomic quantities and the policy rate was described

[14]The monthly time series on logarithmic growth rates of both CCPI and NFP may be obtained from here and here.

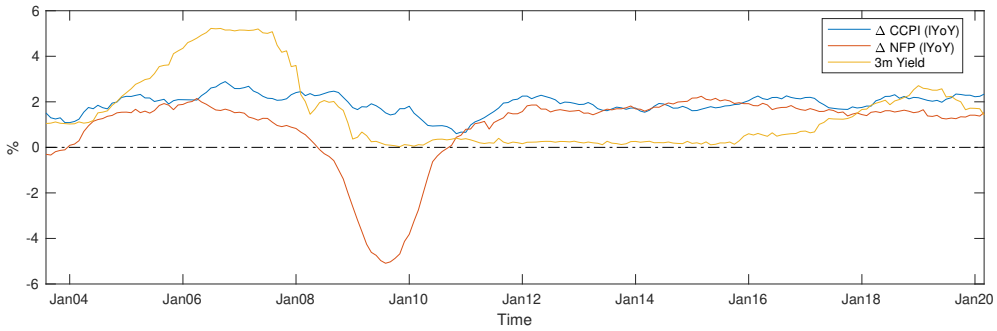

Figure 5: Logarithmic annual growth series on Core Consumer Price Index (CCPI) and NonFarm Payroll (NFP) over the period July 2003 - February 2020 with the empirical short rate from Table 2.

by Taylor's rule of thumb (e.g. see Woodford (2001) for an explanation in greater detail). In contrast, this stylized relation appears to be broken from November 2008 to October 2015 as the overnight rate was bounded by the ZLB (Bernanke (2015)).

5 Results

In this part of the paper, we bring the yield curve models from Section 2 to the data from Section 4. For readability purposes, we categorize our models in two sets. The first set comprises the yields-only specifications, which refer to the models from Sections 2.1, 2.2 and 2.3. The second set consists of the macro-finance specifications, i.e. the models proposed in Sections 2.4 and 2.5, respectively. We analyze the in-sample fits in Section 5.1, where we compare the yields-only specifications in Section 5.1.1 and the macro-finance specifications in Section 5.1.2. Then, we assess the predictive performance of each specification in Section 5.2 with the same categorization.

5.1 In-Sample Analysis

In this Section, we elaborate on the posterior in-sample results using the methods from Section 3. We follow Mönch (2012) and run the sampling routines for 125,000 iterations, where we discard the first 25,000 observations and solely store every 10th sample to obtain near-independent draws. To quantify our results, we seek to minimize the quadratic loss functions concerning the posterior densities, i.e. the reported point estimates correspond to posterior means. Furthermore, we provide posterior standard deviations, 95% Highest Posterior Density Intervals (HPDIs) and single chain convergence diagnostics from Geweke et al. (1992).[15] The tests can not be rejected for most parameters with 95% confidence levels. We additionally assure the MCMC chains to be converged beyond our proposed diagnostic with a visual inspection of traceplots, sampled histograms and Empirical Auto-Correlation Functions (EACFs).

[15] As a result of almost eliminated empirical auto-correlations, the proposed diagnostic converges to a more general equal mean test, which asymptotically follows a standard normal distribution. We follow Geweke et al. (1992) to compare the first 10% with the last 50% of effective random samples.

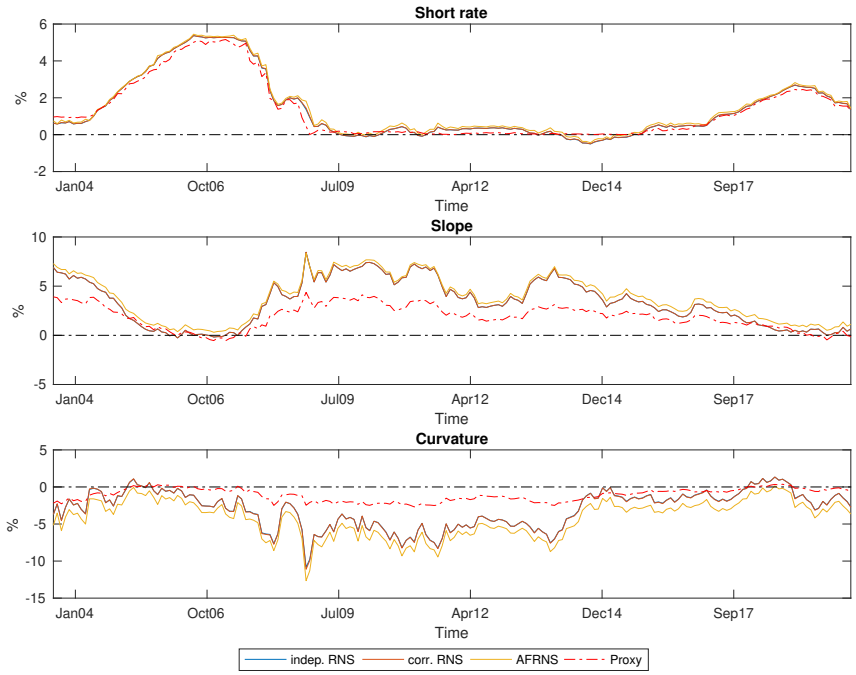

Figure 6: Time series of the latent factors from the (AF)RNS models over the period July 2003 - February 2020. The red-dotted lines correspond to the respective empirical counterparts as mentioned in the second panel of Table 2.

5.1.1 Yields-Only Models

We start our in-sample analysis with an examination of the latent factors from the yields-only specifications. Figure 6 demonstrates that the latent factors of the (AF)RNS models follow the movements of their respective empirical approximates (e.g. see panel B of Table 2). Moreover, we find that the evolution of the three factors from both RNS specifications coincide over time. We observe a slight deviation for the latent factors of the AFRNS model, which is due to the imposed bias correction from the yield-adjustment. Interestingly, the acquired theoretical short rates do not comply with Black (1995) as they fail to respect the ZLB at the announcements of LSAP1 and LSAP2, respectively. Moreover, the found negative spread at the collapse of Lehman Brothers in 2007 provides further support to an inverted yield curve portraying future economic downturn (Estrella & Mishkin (1996); Mendez-Carbajo (2019)).

Several empirical properties of the DNS model persist with the factor rotation from Nyholm (2018) in our empirical exercise. In line with recent advances in literature as Hautsch & Ou (2008) and Koopman et al. (2010), we observe fierce but gradually declining degrees of persistence for both RNS models in Table 5 from Appendix H.1. The short rate is close to being nonstationary for both models, which finds empirical support from Cieslak (2018). Also, Figure 7 provides parametric evidence that underpins the inclusion of both correlated dynamics and correlated disturbances within the RNS framework. The need for correlated factor dynamics stems from the fact that both $F_{1,2}$ and $F_{1,3}$ do not feature zero in their respective 95% HPDI domains, which substantiates the role of slope and curvature in the evolution of the theoretical short rate. Similarly, the requirement for a non-diagonal covariance matrix $\boldsymbol{Q}$ arises from the

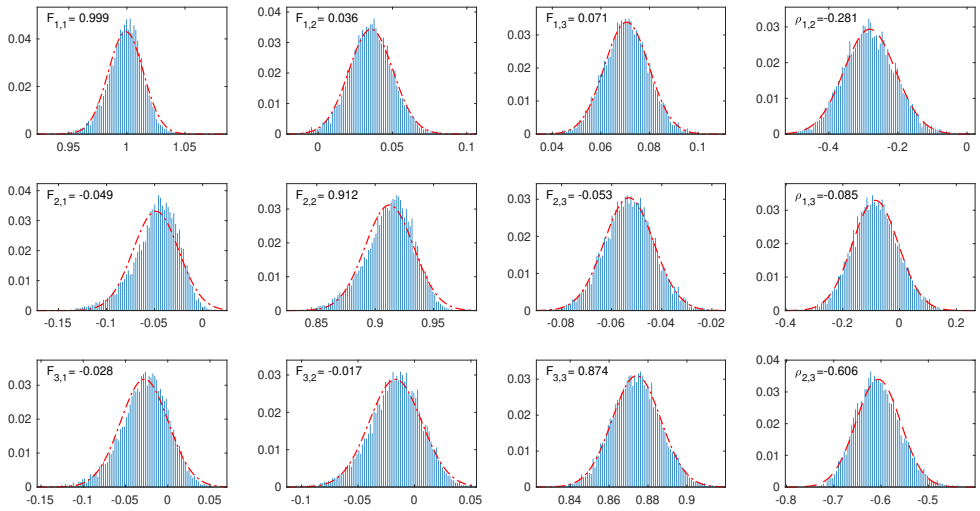

Figure 7: In-sample posterior densities of the state dynamics and disturbance correlations for the correlated factor RNS model, where the values coincide with the posterior means. The red lines correspond to fitted normal distributions with means equal to the respective posterior means and variances equal to the squared respective posterior standard deviations for all parameters considered.

absence of zero in the 95% HPDIs of $\rho_{1,2}$ and $\rho_{2,3}$, respectively.

We present our posterior in-sample results of the AFRNS model in Table 6 from Appendix H.1 while we additionally assure the convergence of the continuous-time parameters with Figures 20-22 from Appendix H.2. Firstly, an inspection of the discretized state dynamics and covariance matrix in the first panel of Table 6 reveals similar values across the independent factor RNS and AFRNS models. More specific, the differences between the acquired state dynamics of both specifications fall within one hundredth as the similarities emerge from AFRNS's unrestricted $\mathbb{P}$-dynamics. Secondly, the evidence that the continuous-time parameters from the AFRNS model have converged is two-fold. On the one hand, we find that all convergence diagnostics from Geweke et al. (1992) can not be rejected with 99% certainty. On the other hand, visual inspection of Figures 20-22 shows traceplots fluctuating around their posterior means, posterior densities aligning with their sampled histograms and near-zero EACFs.

The first panel of Figure 8 shows that the imposition of arbitrage-free dynamics in both the AFRNS and AFNS model is a disruptive factor for the in-sample fit. Precisely, we find near-identical mean fitted yield curves within the set of arbitrage-free specifications and within the set of models that allow for arbitrage opportunities. Comparison between the two groupings shows deviations at both ends. All specifications fail to model the near-zero short-end during the ZLB period as we find overly high modelled values for low maturities (Andreasen & Christensen (2015)). This inability further increases with our arbitrage-free models. At mid-term maturities, the fitted yields follow a near-identical pattern, whereas the behavior between the two groupings diverges at the long-end. Here, the yield-adjustment exerts downward pressure on both AFRNS and AFNS to achieve the desired concavity which RNS and DNS both lack. Nonetheless, the implied flattening appears too rigid as the fitted curves of both arbitrage-free models fall below the observed empirical mean of the ten-year treasuries. This especially holds for the AFRNS

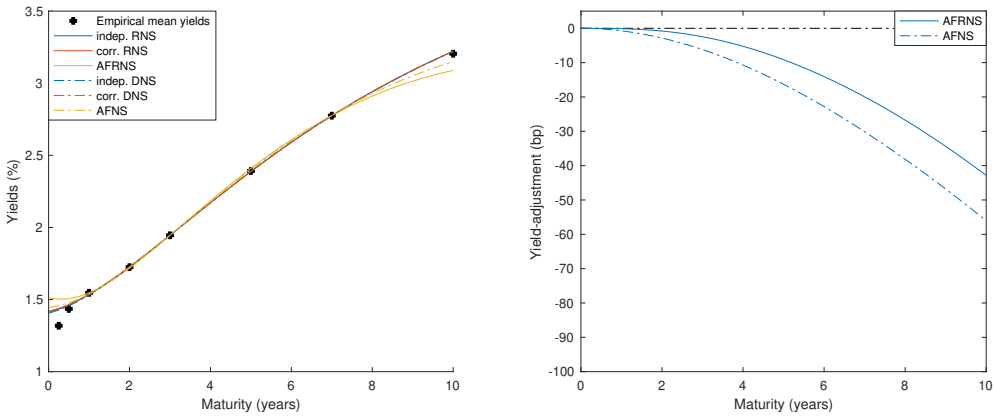

Figure 8: The left partition shows the mean zero-coupon yield curves for all yields-only specifications over the period July 2003 to February 2020. The right partition shows the estimated yield-adjustments of both arbitrage-free Nelson-Siegel models. The estimated parameters in this figure are acquired via the Bayesian estimation procedures from Section 3.

model.

The second panel of Figure 8 shows the yield-adjustments of both the AFRNS model and the AFNS model from Christensen et al. (2011) to have a similar evolution with maturity. More specific, both yield-adjustments are smooth and strictly negative functions that monotonically decrease towards the long-end of the yield curve. The modelled concavity they bring in the first panel of Figure 8 advocates the inability of both standard RNS and DNS to be arbitrage-free (Christensen et al. (2011); Coroneo et al. (2011)). Suppose that one takes a bullish approach towards long-term bonds and seeks to hedge his position with a slightly sooner expiring security. In that case, the yield curve must ultimately slope downwards due to Jensen's inequality. Both RNS and DNS models lack this decelerating feature for higher maturities.

Table 3: In-sample performance of yields-only specifications with our Bayesian approach

	Rotated Nelson-Siegel (RNS)						**Dynamic Nelson-Siegel (DNS)**					
	Independent		**Correlated**		**Arbitrage-Free**		**Independent**		**Correlated**		**Arbitrage-Free**	
Maturity	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
3	-11.52	**21.61**	-12.44	22.27	-18.52	26.63	*-10.94*	21.91	-12.00	21.91	-13.93	23.62
6	-2.64	**9.17**	-3.40	9.45	-7.12	11.85	*-2.16*	9.71	-3.04	9.26	-4.24	10.56
12	1.20	2.43	0.70	1.53	*0.00*	**0.03**	1.52	3.33	0.94	1.98	0.72	3.08
24	-0.10	2.03	-0.28	2.11	0.72	2.31	*-0.00*	1.83	-0.19	2.09	0.21	**1.81**
36	-0.00	0.01	-0.00	0.01	*0.00*	**0.00**	-0.01	0.06	-0.00	0.00	0.00	0.15
60	0.33	2.48	0.41	2.55	-1.78	3.07	0.28	**2.34**	*0.37*	2.52	-0.73	2.56
84	0.00	0.01	*-0.00*	**0.01**	0.01	0.01	0.00	0.05	0.00	0.01	-0.14	1.23
120	-1.63	9.05	-1.82	9.26	11.72	14.57	*-1.51*	**8.69**	-1.73	9.18	6.06	11.14
All	-1.80	**5.85**	-2.10	5.90	-1.87	7.31	*-1.60*	5.99	-1.96	5.87	-1.51	6.77

Note: This table contains the mean residuals (Mean) and Root Mean Squared Errors (RMSEs) in basis points (bp) over the period July 2003 to February 2020 for all yields-only specifications and maturities considered. The maturities are reported in months. The mean residuals are constructed as the difference between the realized yields minus the fitted yield. Italic (Bold) entries correspond to the cross-sectional absolute lowest mean residual (RMSE) for that maturity. Additionally, green cells correspond to lower absolute mean residuals or RMSEs in comparison to Table 9.

In Table 3, we estimate our yields-only specifications with the Bayesian methodologies from Section 3 to compare the average in-sample fits for all maturities. We propose two measures to assess the adequacy of the in-sample fit. Under Mean, we plug-in the estimated parameters to construct the mean residuals as the difference between the realized yields and the fitted

yields. Thus, a value larger (smaller) than zero indicates underestimation (overestimation) on average for that specification and maturity. An italic entry corresponds to the absolute lowest mean residual for that maturity. Under RMSE, we construct the Root Mean Squared Errors (RMSEs) for all maturities and specifications considered. A bold entry indicates the lowest RMSE for that particular maturity. Additionally, we compare the in-sample fit of our Bayesian approach with the in-sample fit from classical methods (e.g. see Appendix C.3). As such, a green entry corresponds to a lower absolute mean residual or RMSE generated with our Bayesian approach (e.g. see Table 9 in Appendix H.1 for the reported criteria of assessing the in-sample fit with classical methods).

A cross-sectional comparison among our yields-only specifications shows competitive in-sample fits in Table 3. We find that the overall RMSEs of the RNS, AFRNS, DNS and AFNS models all lie close with values of 5.85, 5.90, 7.31, 5.99, 5.87 and 6.77, respectively. In line with Christensen & Rudebusch (2016), we find near-zero mean residuals and RSMEs for some maturities. This indicates that that respective model loads on that particular yield. The findings from Figure 8 find quantitative support in Table 3. Specifically, we observe relatively strong negative mean residuals and relatively large RMSEs at short-term maturities, which indicates the inadequacy of all specifications considered to model the near-zero short-end during the ZLB period (Bauer et al. (2014); Christensen & Rudebusch (2015)). This primarily holds for the arbitrage-free models with lower mean residuals and higher RMSEs compared to their statistical counterparts. At the long-end of the yield curve, we find both RNS and DNS unable to model the desired concavity as we find negative mean residuals and relatively large RMSEs. In contrast, both AFRNS and AFNS exert too much downward pressure on the long-end with high positive mean residuals and larger RMSEs. Also, we find the compression at the long-end to be more rigid with the AFRNS model.

Table 3 also shows that the greater flexibility from correlated factor dynamics does not necessarily result in an improved in-sample fit. We observe that the overall RMSE of the correlated factor RNS model is 0.05 basis points higher than the overall RMSE of its less flexible counterpart, which contradicts the evidence from Figure 7. The opposite holds for the DNS model, i.e. we find an overall reduction of 0.12 basis points in RMSE to arise from the allowance for correlated dynamics. Yet, the observed differences between independent factor dynamics and correlated factor dynamics are small within our in-sample study.

Next, we shift focus in Table 3 to compare the in-sample results from the Bayesian method with the classical Nelder-Mead simplex method (e.g. see Table 9). First, we find that Bayesian estimation yields promising results for the statistical specifications as we find the overall RMSEs of both RNS and DNS models to be highlighted in green. This surplus of explanatory power largely stems from the acquired in-sample fits at the short-end, i.e. we find slightly lower RMSEs for low maturities and approximately equal fits at mid-term and long-term maturities with the Bayesian approach. In contrast, we find less in-sample adequacy to arise from Bayesian estimation of arbitrage-free specifications. Comparison between the overall RMSEs of the arbitrage-free models in Tables 3 and 9 yields a 0.91 basis point increase (+12.5%) for the AFRNS model and a 0.27 basis point increase (+4.15%) for the AFNS model. Still, the differences

between both estimation methods fall within the range of one basis point and may therefore be considered small.

5.1.2 Macro-Finance Models

Many empirical findings on the statistical models from Section 5.1.1 remain with the inclusion of our proposed macro-finance factors. Firstly, Figure 23 from Appendix H.2 shows that the latent factors of the macro-finance models closely follow the evolution of their respective empirical approximates (e.g. see panel B of Table 2). Secondly, we find the high degree of persistence on the latent factors to prevail with the statistical macro-finance models, i.e. Table 7 in Appendix H.1 shows the autoregressive coefficients of both MF-RNS models to have values above 0.9. Thirdly, we find Bayesian support for a correlated variant of the MF-RNS model in Figure 24 as $K_{1,3}$, $\rho_{1,2}$ and $\rho_{2,3}$ do not feature zero in their respective 95% HPDIs. This implies the need for a non-diagonal structure on both $\boldsymbol{K_1}$ and $\boldsymbol{Q}$, respectively (e.g. see Equation (30)). Lastly, the coefficients of the independent factor MF-RNS model from Table 7 and the discretized state dynamics from the MF-AFRNS model in the second panel of Table 8 resemble due to MF-AFRNS's unrestricted $\mathbb{P}$-dynamics.

We find the deviating fit at the cross-sectional tails for no-arbitrage specifications to remain with the macro-finance models. Inspection of the first panel from Figure 9 demonstrates that the inclusion of exogenous factors does not result in a better mean fit for the yields from the H.15 series (e.g. see Section 4.1). Moreover, we recognize the same distinction between statistical and no-arbitrage specifications as in Section 5.1.1, i.e. the arbitrage-free models tend to increasingly overestimate the short-end of the yield curve. Both groupings show similar mean fitted yield curves at mid-term maturities, whereas we find the same diverging behavior from Section 5.1.1 at long-term maturities. In line with their yields-only counterparts, both variants of MF-RNS and MF-DNS lack the ability to model the desired downward pressure at the long-end of the yield curve. In contrast, the negative tilt of both the MF-AFRNS and MF-AFNS model appears too rigid for larger maturities. Yet, we find that the difference in average fit with the ten-year treasuries between the MF-AFRNS and MF-AFNS models is smaller than the found difference between the AFRNS and AFNS models.

Table 4 quantifies the findings from the previous paragraph and compares the in-sample fits from our macro-finance specifications. We find that the inclusion of exogenous factors does not result in overall in-sample improvements for statistical models. Specifically, we obtain a difference in overall RMSEs of +0.02 and +0.07 basis points for the MF-RNS models and +0.00 and +0.07 for the MF-DNS models when compared to the findings from Table 3. Similar to Section 5.1.1, we find the macro-finance specifications to be loaded on maturities with near-zero mean residuals and RMSEs (Christensen & Rudebusch (2016)). We find opposing evidence for the arbitrage-free specifications, i.e. we find the RMSE of the MF-AFRNS model to decrease with 0.20 basis points and the RMSE of the MF-AFNS model to increase with 0.20 basis points compared to their yields-only counterparts. More general, the differences between our yields-only specifications and our macro-finance specifications in terms of in-sample fit appear small. Besides, we find the ambiguity between independent factors and correlated factors to

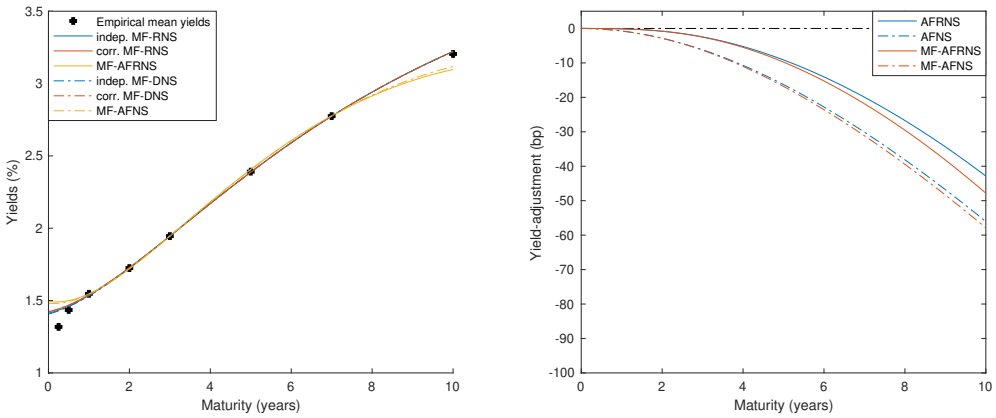

Figure 9: The left partition shows the mean zero-coupon yield curves for all macro-finance specifications over the period July 2003 to February 2020. The right partition shows the estimated yield-adjustments of both arbitrage-free Nelson-Siegel models. The estimated parameters in this figure are acquired via the Bayesian estimation procedures from Section 3.

Table 4: RMSE (in basis points) for macro-finance specifications with our Bayesian approach

	Macro-Finance Rotated Nelson-Siegel (MF-RNS)						**Macro-Finance Dynamic Nelson-Siegel (MF-DNS)**					
	Independent		**Correlated**		**Arbitrage-Free**		**Independent**		**Correlated**		**Arbitrage-Free**	
Maturity	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
3	-11.57	**21.72**	-12.42	22.33	-17.27	25.65	*-11.06*	21.98	-12.42	22.33	-16.37	24.78
6	-2.68	**9.25**	-3.38	9.54	-6.44	11.32	*-2.26*	9.72	-3.39	9.54	-5.81	10.72
12	1.18	2.43	0.71	1.84	*0.03*	**0.16**	1.46	3.26	0.71	1.83	0.28	1.32
24	-0.11	2.02	-0.27	2.07	0.48	2.29	*-0.02*	**1.82**	-0.27	2.07	0.44	2.07
36	-0.00	0.01	*0.00*	**0.00**	-0.01	0.01	-0.01	0.04	0.00	0.01	-0.00	0.02
60	0.33	2.47	0.41	2.52	-1.50	3.01	*0.29*	**2.34**	0.41	2.52	-1.26	2.80
84	*0.00*	**0.01**	-0.00	0.01	-0.05	0.05	0.00	0.06	-0.01	0.01	-0.03	1.12
120	-1.64	9.04	-1.81	9.18	10.90	14.37	*-1.53*	**8.70**	-1.81	9.18	9.05	12.91
All	-1.81	**5.87**	-2.09	5.94	-1.73	7.11	*-1.64*	5.99	-2.10	5.94	-1.71	6.97

Note: This table contains the mean residuals (Mean) and Root Mean Squared Errors (RMSEs) in basis points (bp) over the period July 2003 to February 2020 for all macro-finance specifications and maturities considered. The maturities are reported in months. The mean residuals are constructed as the difference between the realized yields minus the fitted yield. Italic (Bold) entries correspond to the cross-sectional absolute lowest mean residual (RMSE) for that maturity. Additionally, blue cells correspond to lower absolute mean residuals or RMSEs in comparison to Table 3.

prevail with the macro-finance specifications. For the MF-RNS model we observe a slight decline in overall performance with increasing RMSEs, whereas we find a minimal overall increase in performance for the MF-DNS model. Yet, the differences between independent factors and correlated factors continue to be small with our macro-finance specifications.

Although the inclusion of exogenous factors does not necessarily yield in-sample improvements for all specifications, it does provide a greater understanding of the modelled bi-directional dynamics between the latent factors and the observed series in the MF-(AF)RNS models. We follow Diebold & Li (2006) and provide the Forecast Error Impulse Response Functions (FEIRFs) in Figure 10 of the VAR(1) process on the state equations.[16] We distinguish the behavior of four groupings: exogenous responses to exogenous shocks, exogenous responses to term structure shocks, term structure responses to exogenous shocks and term structure

[16]Within econometric literature, there appears to be three popular options of impulse response functions, i.e. Forecast Error Impulse Response Functions (FEIRFs), Orthogonal Impulse Response Functions (OIRFs) and Generalized Impulse Response Functions (GIRFs) (e.g see Koop et al. (1996) and Pesaran & Shin (1998)). Due to its popularity, uniform interpretation and robustness to the order of variables we opt for FEIF.

responses to term structure shocks.[17]

Preliminary inspection demonstrates that the impulse responses are relatively robust to the exclusion of arbitrage opportunities and the allowance for correlated factors. The plotted lines for the independent factor MF-RNS and MF-AFRNS models follow similar silhouettes. In contrast, those for the correlated factor MF-RNS model slightly deviate in magnitude compared to the former two. We find the most substantial deviations in the off-diagonal responses from partition $\boldsymbol{K_1}$ as the independent factor models restrict these coefficients to be zero.

Interpretation-wise, the impulse responses refer to reactions of the variable shown on the left of a respective row to ceteris paribus unit (e.g. percentage) shocks of the variable presented above that column in Figure 10. For example, the fourth entry of the first row corresponds to the impulse response of the modelled theoretical short rate to a unit shock on annual CCPI growth. Here, we observe a prompt negative adjustment of the short rate to a unit shock of inflation with a further decline over the next 30 months. An inverse response is to be expected as the Fed traditionally seeks to combat rising inflation with incremental changes on the overnight rate. Within our in-sample analysis, however, the overnight rate possesses no more slack, and hence the Taylor Rule appears to be broken (Bernanke (2015)).

Analogue to the interpretation of our example, we first consider the within-set responses of the macro-finance factors. All show great persistence as shocks gradually fade out. In confirmation with Phillips (1958), inflation shows clear positive adaptation to shocks on labour supply. Its effectiveness slowly dissipates towards longer-horizons as a sequential decrease of the MF in response to declining NFP shows the Fed's reaction to achieve its inflation target via QE's widening purpose (Blinder et al. (2010)). As a unit shock of MF has minimal impact on inflation, the effectiveness of LSAPs for other aims than its original goals also appears prone to their magnitudes (Solow et al. (1987); Blinder (2000)). Indeed, this shows the alternated landscape of interest rates where the stimulating abilities of the traditional FFR no longer prevail. Instead, the Fed turns the knobs on the MF of ten-year treasury yields, among other things (Diebold et al. (2006)).

Interaction with the term structure provides new insights into the evolution of the macro-finance factors. In line with traditional findings from Rudebusch & Svensson (1999), we find endogenous shocks on the modelled theoretical short rate to negatively coerce future development of inflation and NFP. Incremental shocks on the first latent factor indicate the balance sheet normalization program with a sequential decrease of the MF (Bonis et al. (2017)). The same principle applies to shocks on the slope factor, i.e. the Fed seeks to loosen its policy grip as a steepened slope indicates future macroeconomic prosperity (Estrella & Mishkin (1996)). Similarly, future economic upswing with the growth of NFP and CCPI arises from a unit shock on the slope factor. However, an initial short-term decline emerges from the model's noncontemporaneous nature.[18]

[17]The behavior of these groupings is explicitly modelled in Tables 7 and 8 in partitions $\boldsymbol{K_4}$, $\boldsymbol{K_3}$, $\boldsymbol{K_2}$ and $\boldsymbol{K_1}$ respectively (see Sections 2.4 and 2.5).

[18]In our studies we implement the exogenous factors as annual growth rates. As a direct consequence, the monthly values of the latent factors may have limited impact on the exogenous factors as they do not affect the first eleven months of the growth figure.

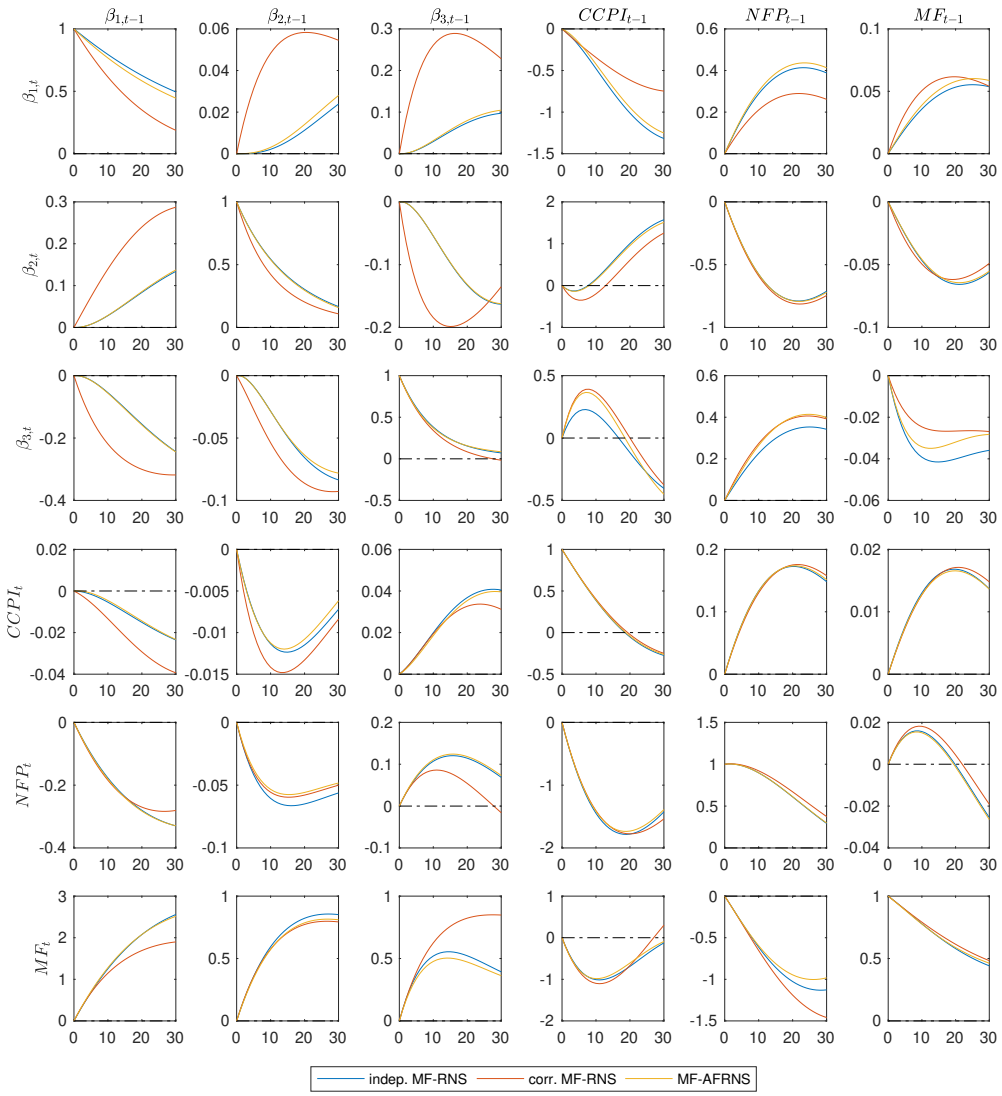

Figure 10: Forecast Error Impulse Response Functions (FEIFs) for mixed latent and exogenous factor specifications. Each partition coincides with the affection of the variable left to the respective row as a result to a unit shock to the variable presented above that respective column. The ordering of the $(J+P)^2 = 36$ partitions coincides with matrices $\boldsymbol{K}$ in Tables 7 and 8.

Next, we examine term structure responses to shocks on the macro-finance factors. As mentioned before, our in-sample analysis suggests that the Taylor Rule is broken since causal incremental changes to the short rate from positive shocks on inflation no longer last. Increments of the term structure's short-end do emerge from shocks on NFP, but they do not necessarily result in increased inflation. In line with Diebold & Li (2006), we find similar responses to the slope factor from shocks on the exogenous factors, i.e. an increase of the MF directly exerts downward pressure to clock-wise rotate the curve.[19] Besides QE's original purpose to negatively coerce the long-end via the scarcity channel from Modigliani & Sutch (1966), a broad surplus of liquidity arising from LSAPs adversely adjusts future evolution of the short rate. Apart from escalating money supply, the signalling channel from Bauer & Rudebusch (2013) further strengthens downward pressure on the short-end of the term structure.

Lastly, we consider term structure responses to lagged term structure shocks. All three exhibit subordinate but declining persistence from the first to the third latent factor. The modelled shadow short rate shows positive adaptation to unit shocks on the slope factor, indicating future improvement of the macroeconomic climate (Estrella & Mishkin (1996)). Indeed, the opposite causal effect also holds as positive shocks on the short rate indicate escape from the ZLB. Furthermore, we find shocks on the curvature factor to flatten the yield curve and possibly predict future economic downturn (Mönch (2012)).

5.2 Out-of-Sample Performance

In addition to the in-sample fit, predictive performance has often been ought to resemble modelling adequacy of term structure specifications (Ang & Piazzesi (2003); Hautsch & Ou (2008); Mönch (2012)). Advantages arising from an increasing amount of parameters often no longer prevail in an out-of-sample exercise as the greater flexibility could be prone to overfitting (Diebold & Li (2006); Çakmaklı (2020)). Thus, we extend our research questions towards an out-of-sample exercise and follow the order from Section 3.1. First of all, does the statistical support for cross-sectional latent factor dynamics yield better forecasts? Secondly, how does the inclusion of no-arbitrage affect predictive ability? Thirdly, does the factor rotation from Nyholm (2018) produce better out-of-sample performance? Lastly, how does our Bayesian estimation approach perform compared to classical estimation methods in a predictive exercise? We answer these questions in Section 5.2.1. We subsequently extend the first three questions to our macro-finance models, whereas we additionally investigate the difference in predictive performance between our yields-only and macro-finance specifications.

We construct h-step ahead forecasts with $h \in \{1, 3, 6, 12\}$ in months using simulation to involve both parameter- and forecast uncertainty (e.g. see Section 3.2 for an explanation in greater detail). Our prediction sample comprises fifty monthly observations, i.e. it starts in January 2016 and ends in February 2020. We generate posterior predictive densities from ten-year rolling windows with an identical sampling procedure as to our in-sample exercise. Hence, the point forecasts correspond to the posterior means of these predictive posterior

[19] The signs mentioned by Diebold et al. (2006) are the opposite to what we find, but the interpretation remains equivalent as the slope switches sign for RNS compared to DNS (Frankel & Lown (1994); Nyholm (2018))

densities. We evaluate the predictive accuracy of our discussed specifications via the difference in Root Mean Squared Forecast Errors (RMSFEs).[20] Additionally, we isolate each of the research questions via the assessment of relative RMSFEs. A value smaller (larger) than one indicates that the first-mentioned model in that partition's title performs better for that horizon and maturity.

Tables 10 and 11 in Appendix H.1 provide the RMSFEs for each specification, maturity and forecast horizon considered. With Table 12 in Appendix H.1, we also grant the RSMFEs for our yields-only models acquired via classical estimation. Additionally, Figures 25-28 in Appendix H.2 show the forecasts for all horizons, maturities and RNS specifications examined. The paths for short-term horizons, i.e. one-month-ahead and one-quarter-ahead, closely follow the evolution of each observed underlying cross-sectional component with RMSFEs often within the range of 25 basis points. However, incremental deviations from the observed data emerge at larger forecast horizons. This is due to the announced balance sheet normalization program by the Fed in November 2017, where the considered rolling windows for larger forecast horizons base their forecasts on a persistent near-zero outlook. Similarly motivated, most specifications fail to forecast the falling rates in 2019 at larger horizons as their parameters were estimated using rolling windows with relatively optimistic rates.

5.2.1 Yields-Only Models

First, we address the difference in predictive performance between independent factors and correlated factors within the yields-only Nelson-Siegel environment. In both partitions of Figure 11, we observe two recurring patterns. Firstly, we observe a bell-shaped silhouette per forecast horizon, where the relative performance of the correlated variants first increases and then gradually decreases towards the long-end of the curve. Secondly, we find the relative performance of both independent factor RNS and DNS to grow with the forecast horizon, i.e. the intercepts of the bell-shaped patterns decrease with expanding forecast horizons.[21] For forecast horizons up to six-months-ahead, we find the correlated factor variants of both Nelson-Siegel specifications to yield additional performance. For larger maturities and the one-year-ahead forecasts the opposite holds, i.e. we see consistent improvement in forecast ability of independent models for ten-year maturities and the largest horizon. These findings extend the clash between both variants of the DNS model from Christensen et al. (2011) to both bases in the post-crisis era.

Figure 12 shows contrasting consequences to arise from the imposition of no-arbitrage dynamics to RNS and DNS, respectively. In the left partition, we find that the independent factor RNS model largely exceeds the AFRNS model in terms of predictive performance. For example, we find that the independent factor RNS model has 62% more accurate predictions for the one-year-ahead forecasts of the ten-year treasuries on average. This indicates that the preclusion of arbitrage opportunities within the AFRNS framework does not result in better forecasts,

[20] In a Bayesian context, Bayes factors or marginal likelihoods often function as a toolkit for model comparison. As the former does not allow for diffuse prior comparisons and the second yields no straightforward computation in a state space setting, we resort to RMS(F)Es.

[21] The term intercept is mathematically invalid here, as the shortest yield considered has a maturity larger than zero months. Yet, it provides greater comprehension on the difference in forecast performance between both methods.

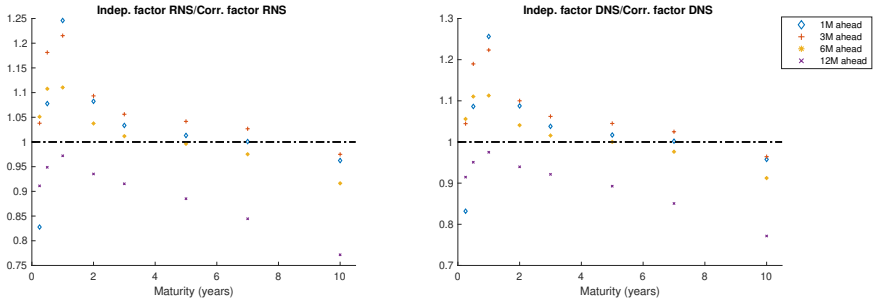

Figure 11: Relative RMSFEs to assess the difference in predictive performance between independent factors and correlated factors for our yields-only specifications.

which finds support with Nyholm & Vidova-Koleva (2012) and Diebold & Rudebusch (2013). Contrarily, we find Bayesian estimation of the AFNS model to near-consistently dominate the independent factor DNS model with relative RMSFEs well below one in the right partition of Figure 12. This finds empirical backing from Christensen et al. (2011) as they also show an improvement of forecast performance to emerge from the freedom of arbitrage in AFNS. Despite their differences, both partitions have one commonality. We find a recurring pattern that increases towards the long-end of the term structure. This contradicts the same pair-wise comparison from Christensen et al. (2011), who find the most substantial relative performance of the AFNS model at higher maturities. This may be due to the different sample periods considered, i.e. the inferred yield-adjustment at higher maturities may be too harsh during the ZLB period.

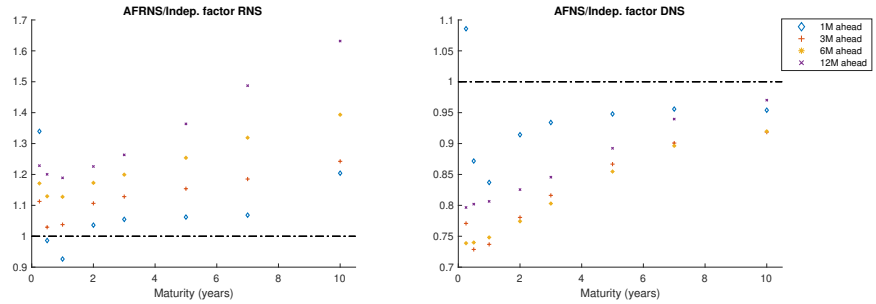

Figure 12: Relative RMSFEs to assess the difference in predictive performance between arbitrage-free and statistical yields-only specifications.

Next, we observe the predictive difference between the traditional Nelson-Siegel factors and the adjusted base from Nyholm (2018) in Figure 13. We find that the observational equivalence between RNS and DNS largely remains within our out-of-sample study (Nyholm (2018)), i.e. we find relative RMSFEs within the range of 0.99 and 1.012 in the first two partitions of Figure 12. Still, we observe that the predictive power of the factor rotation from Nyholm (2018) decreases along the forecast horizon. This near-equivalence in out-of-sample ability ends with the imposition of arbitrage-free dynamics. Specifically, we find all relative RMSFEs to lie above one in the third partition of Figure 13. Hence, we find consistent improvement in forecast performance when we consider the AFNS model over the AFRNS model. Moreover, we find the difference in forecast performance between both arbitrage-free specifications to increase with the forecast horizon considered. This resolves to the formulation of two hypotheses, i.e. the AFNS

model provides a better way to incorporate no-arbitrage dynamics within the Nelson-Siegel environment and/or the performance of the AFRNS model depends on the estimation method considered.

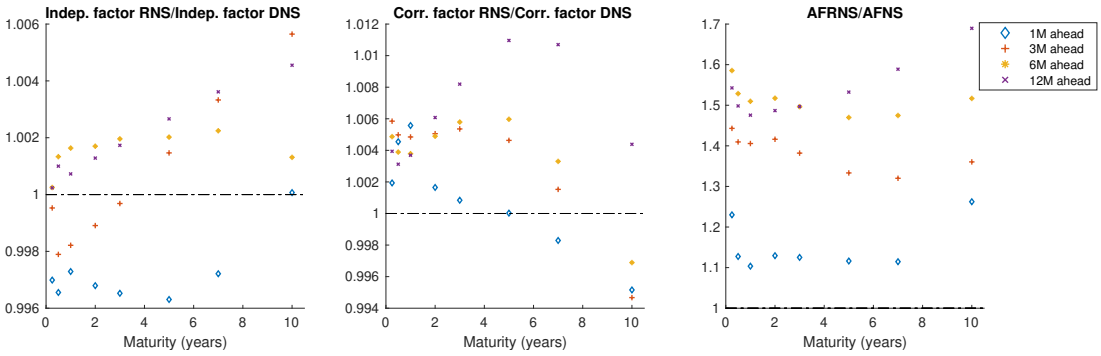

Figure 13: Relative RMSFEs to assess the difference in predictive performance between the factor rotation from Nyholm (2018) and the traditional Nelson-Siegel base from Diebold et al. (2006) for our yields-only specifications.

In Figure 14, we quantitatively assess the influence of our Bayesian method, as we directly compare the constructed forecasts with the classical Nelder-Mead simplex method (e.g. see Appendix C.3). Two relations arise from visual inspection of the left four partitions. Firstly, we observe that the classical method better copes with the short-end of the yield curve, apart from the one-month-ahead forecasts. Secondly, we find that the relative performance of our Bayesian method for statistical models increases along the maturity dimension. These findings find further support in Table 10, i.e. we find green entries at one-month-ahead forecasts of the short-end and we find green entries for one-year-ahead forecasts of the long-end from our statistical specifications.

The last column of Figure 14 provides conflicting evidence for the Bayesian estimation approach in a no-arbitrage setting. The lower left partition shows our proposed estimation method to be an effective means to incorporate no-arbitrage dynamics within the Nelson-Siegel environment. We find that most relative RMSFEs are valued below one, such that our newly constructed Bayesian approach outperforms the classical method on most horizons and maturities considered. Yet, the upper right partition tempers our enthusiasm. Specifically, we find a substantial improvement in predictive adequacy to arise from the estimation of the AFRNS model with the classical method as all relative RMSFEs are valued above one.

Consequentially, we seek to identify the root cause of the contrasting predictive performance between both arbitrage-free specifications. We demonstrate via Figure 29 in Appendix H.2 that the predictive failure of the AFRNS model in our out-of-sample study is not model-specific. Precisely, we extend the clashes that involve the AFRNS model from Figures 12 and 13 to the classical approach. The left partition of Figure 29 shows that the imposition of no-arbitrage dynamics can enhance performance within the RNS framework. Additionally, the right partition shows that the predictive performances of both arbitrage-free specifications are competitive in maximum likelihood setting.

Hence, we investigate the limitations of our Bayesian method using our empirical out-of-sample study. Figure 30 in Appendix H.2 shows that the techniques from Section 3

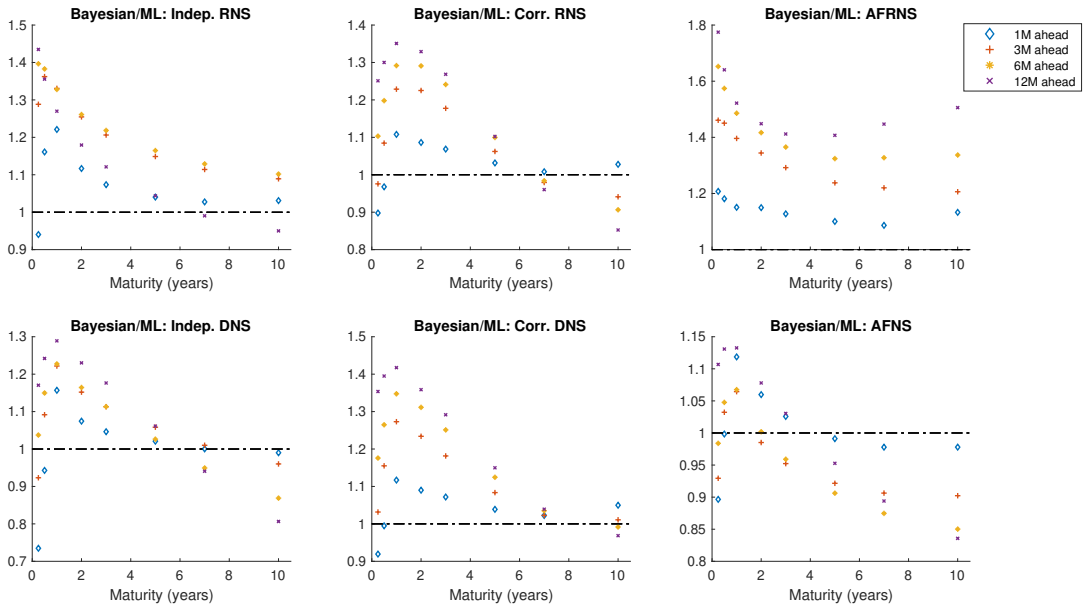

Figure 14: Relative RMSFEs to assess the difference in predictive performance between our proposed Bayesian estimation method and classical estimation methods.

offer limited statistical freedom for the estimation of the continuous-time drift processes from Equation (23). We find that the acquired posterior results per rolling window from the Bayesian method differ slightly from the imposed respective prior means. In contrast, we observe that the unconditional means obtained via maximum likelihood are much more volatile. This indicates that the data-driven prior from Diebold et al. (2008) may be too restrictive for the AFRNS model and that more flexibility may be desired. Besides, Christensen & Rudebusch (2015) demonstrate that forecasts for longer horizons quickly converge to their unconditional means, which also explains the diverging performance at longer horizons between both arbitrage-free specifications in the right partition of Figure 12. The second limitation of our method lies in the fixation of λ via the preliminary two-step method. Yet, Nelson & Siegel (1987) mention the value of λ to have little impact on the results and Çakmaklı (2020) show destabilization of the MCMC chains to emerge from freely sampling λ. Hence, we postulate the restrictive priors on the unconditional means of the AFRNS model to be the driving forces behind its lacking forecast performance.

5.2.2 Macro-Finance Models

Table 11 and Figure 15 address the predictive influence of including inflation, economic activity and our suggested supply factor in the term structure specifications. Two relations show up after visual inspection of the left four partitions, although the upper left partition forms an exception. Firstly, we find a near-consistent surplus of predictive performance to arise for short-term and mid-term maturities with our macro-finance models, i.e. most relative RMSFEs for maturities between six months and seven years are valued below one in Figure 15. Hence, the implementation of inflation, economic activity and a monetary policy instrument appears crucial

for those yields (De Pooter et al. (2007); Yu & Zivot (2011)). The one-step-ahead forecasts on the yields with three months to maturity are exempted from this pattern as the inclusion of macro-finance factors possibly distorts the persistent nature of the short-end during the ZLB period. The second stylized relation emerges from increasing relative forecast performances along the horizon for most macro-finance specifications considered. For the arbitrage-free specifications, we obtain an opposing influence of the proposed observed factors. On the one hand, we find near-uniform improvement in forecast performance of the MF-AFRNS model, which may also be due to its lesser dependence on the unconditional means. On the other hand, we find the predictive performance of the MF-AFNS model to be near-uniformly lower than its yields-only counterpart.

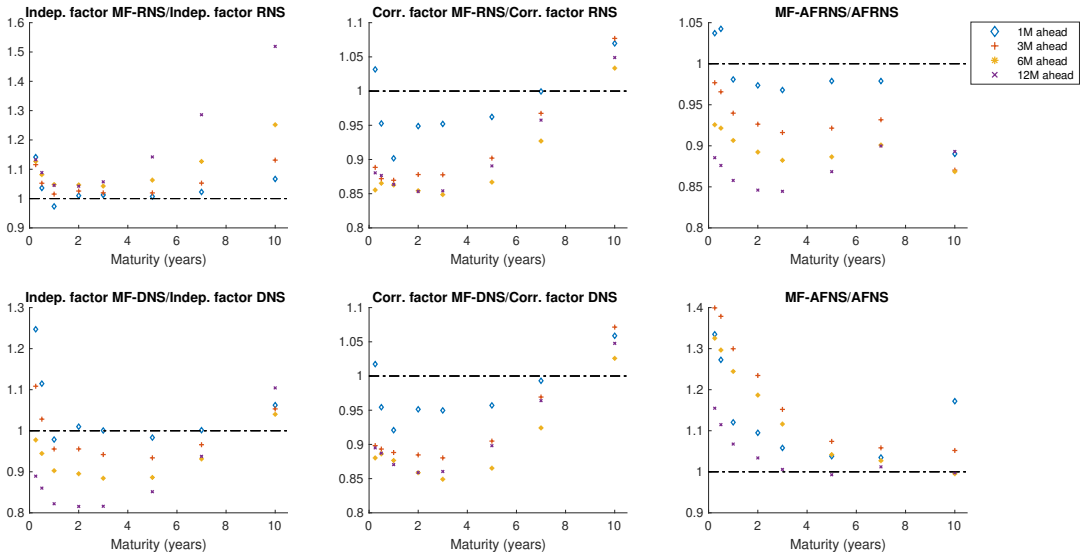

Figure 15: Relative RMSFEs to assess the difference in predictive performance between macro-finance specifications and yields-only specifications.

Next, we portray the difference in predictive performance between independent factor macro-finance specifications and their more flexible counterparts in Figure 16. At first glance, we observe a similar bell-shaped pattern from Figure 11 per horizon considered. Insensitive to the selected Nelson-Siegel base, we find correlated factor models to near-consistently dominate for maturities up to seven years (e.g. only the one-year-ahead forecasts from the MF-DNS model are exempted from this pattern). In line with Figure 11, we see a gradual improvement in the relative predictive performance of independent factor macro-finance models towards the long-end of the curve. For our yields-only specifications, this resulted in exceeding ability of the independent factor variants for yields with seven and ten years to maturity, respectively. Yet, Figure 16 shows this outperformance to only endure for the MF-DNS model.

The predictive abilities of our macro-finance specifications are relatively robust to the imposition of no-arbitrage dynamics. In Figure 17, we observe that the freedom of arbitrage provides an advantage for both MF-AFRNS and MF-AFNS in predicting the long-end of the yield curve, which coincides with the findings from Christensen et al. (2011). We find the relative RMSFEs to lie within the range of 0.95-1.05 for all other maturities, which indicates

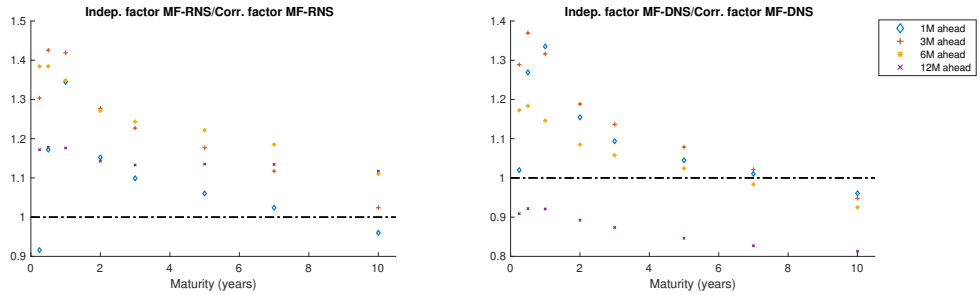

Figure 16: Relative RMSFEs to assess the difference in predictive performance between independent factors and correlated factors with macro-finance specifications.

near-equivalent performance between statistical macro-finance specifications and arbitrage-free macro-finance specifications in our empirical exercise.

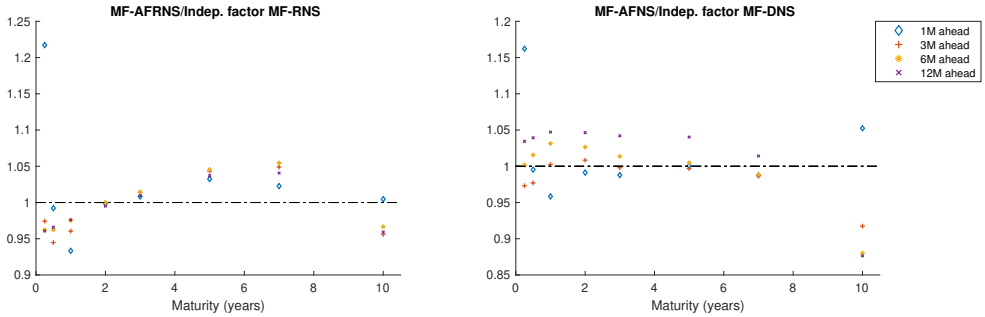

Figure 17: Relative RMSFEs to assess the difference in predictive performance between statistical and arbitrage-free macro-finance specifications.

Next, Figure 18 shows contradictory evidence on the possible equivalence between MF-RNS and MF-DNS in our empirical exercise. The first partition shows that the performance between both statistical bases diverges. Although we obtain comparable one-step-ahead forecasts, we find that the relative performance of the independent factor MF-DNS model uniformly increases with the forecast horizon considered. In contrast, we find near-identical performance between both Nelson-Siegel bases when we consider correlated factor dynamics in our macro-finance specifications. More specific, we find relative RMSFEs valued between 0.975 and 1.02, which largely coincides with the yields-only findings from Figure 13. We find that the performance between both arbitrage-free specifications has converged compared to their yields-only counterparts. If we compare the last partitions of Figures 13 and 18, we see a consistent lowering of all relative RMSFEs considered. This finds support in Figure 15, i.e. the inclusion of macro-finance factors fares well with the MF-AFRNS model and lowers the predictive performance of the MF-AFNS model.

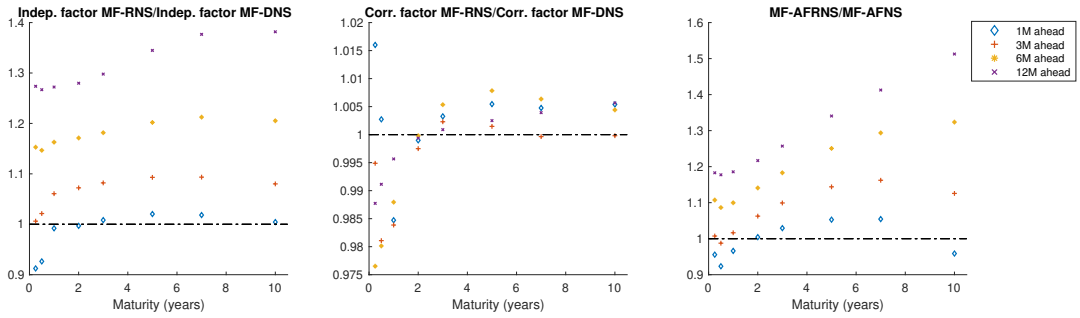

Figure 18: Relative RMSFEs to assess the difference in predictive performance between the factor rotation from Nyholm (2018) and the traditional Nelson-Siegel base from Diebold et al. (2006) for our macro-finance specifications.

6 Conclusion

Both modelling and forecasting of the yield curve are fundamental to trading, asset allocation and risk management. An improved understanding of the underlying term structure dynamics results in improved execution of the tasks as mentioned earlier. As Duffee (2002) shows that the traditional class of ATSMs fails to provide adequate predictive performance, we focused on three developments that emerged in literature afterwards.

Firstly, we resorted to the statistical power from the Nelson-Siegel framework. We departed from the factor rotation by Nyholm (2018), which is observational equivalent to DNS and explicitly identifies the short rate. The RNS model and the class of ATSMs have their advantages for complementary reasons. The class of ATSMs provides the theoretical no-arbitrage restrictions but lacks empirical success. In contrast, the RNS model fits and predicts the yield curve to a greater extend, yet fails to meet the arbitrage-free conditions (Björk & Christensen (1999); Filipović (1999)). Hence, we combined RNS's factor-interpretation with the arbitrage-free framework from Duffie & Kan (1996) in the AFRNS model. The AFRNS model combines the advantages that come with the Nelson-Siegel environment and the class of ATSMSs. At the same time, it becomes more appealing from a policymaker's perspective since central banks traditionally manipulate the short rate to fulfil their mandates. Moreover, we found that the AFRNS model closely relates to the AFNS model from Christensen et al. (2011) as both are restricted versions of the canonical three-factor affine structure from Dai & Singleton (2000). Besides, we have shown that the yield-adjustments of both specifications are linearly interchangeable.

Secondly, we departed from Diebold & Li (2006) and Nyholm (2018) to specify a yield curve specification that incorporates the rotated factors (short rate, slope, curvature) and observed series (inflation, economic activity, monetary policy stance) in MF-RNS. Most papers bring the federal funds rate as a monetary policy variable to the yield curve. However, it is limited in the ZLB period, such that we resorted to the marketable fraction of long-term yields instead. Its influence on the yield curve stems from theoretical sources, i.e. the scarcity channel from Modigliani & Sutch (1966) and the preferred habit view from d'Amico et al. (2012), while several event studies demonstrate its impact to remain in practice. The state space representations of these macro-finance specifications greatly ease estimation and allow to identify relations between

the latent factors and the observed series via impulse response functions. Nonetheless, the MF-RNS model lacks theoretical rigour as it does not comply with the arbitrage-free conditions. Hence, we followed the AFNS-Macro specification from Ullah (2016) and combined no-arbitrage dynamics, the inclusion of exogenous series and the rotated factors from Nyholm (2018) in the MF-AFRNS model.

Thirdly, we extended on the greater interest shown by econometric literature in Bayesian estimation methodologies. Bayesian analysis has several statistical advantages, i.e. it does not suffer from convergence to local maxima (Kim & Wright (2005)) or costly initialization routines via simulation studies (Ullah (2016)). There are numerous advances in Bayesian modelling to estimate models within the Nelson-Siegel environment. However, no advances have been made to allow for no-arbitrage dynamics in a Bayesian domain. Hence, we composed a method that exploits AF(R)NS's unrestricted $\mathbb{P}$-dynamics to reversely tie the Nelson-Siegel specifications to the arbitrage-free framework from Duffie & Kan (1996).

Additionally, we brought our methodology to empirical data via two analyses, i.e. an in-sample study and a forecasting exercise. We used a monthly series of end-of-month U.S. zero-coupon bond yields from July 2003 to February 2020 to answer the five research questions from Section 1. With our in-sample analysis, we empirically showed that the average fit of the yield curve is robust to each of the proposed modifications from the research questions. More specific, we found that the maximal difference in mean in-sample errors was 1.5 basis points among all modelled aspects. Nevertheless, we found some deviations between the proposed specifications when we segmented along the maturity dimension. Although all specifications failed to model the low persistent short-end during the ZLB period, we found this to primarily hold for the arbitrage-free models from the second research question. On the other end, we found that the arbitrage-free specifications were able to model the concavity for high maturities. However, the implied flattening from these models was too rigid on average at the long-end of the yield curve. This prevailed with both Nelson-Siegel bases and estimation methods from the third and fourth research questions, respectively, which contributes to the statement that arbitrage-free specifications may be overly restrictive to model the term structure's long-end during the ZLB period. Lastly, we found that the inclusion of our observed factors from the fifth research question did not improve the average fit. Nonetheless, they have shown to offer a greater comprehension of the bi-directional relations between the macroeconomy, the yield curve and the decisions of a monetary policymaker.

We found more varying results with our out-of-sample analysis. Regarding the first research question, we observed that correlated factor models yield predictive improvement on the short-end of the yield curve. However, the advantage seemingly faded towards the long-end of the term structure. We found that the imposition of no-arbitrage dynamics from the second research question had inconsistent effects in terms of predictive performance. For the AFNS model, we observed a near-uniform increase in forecast ability relative to its no-arbitrage counterpart. This further strengthened with our newly constructed Bayesian method. For the AFRNS model, we found the opposite to hold. On the one hand, we found that the AFRNS model did offer greater predictive performance relative to its statistical counterpart within a classical estimation method.

On the other hand, this advantage vanished with our Bayesian method. Hence, we found that the imposed priors in our Bayesian method were too binding for the AFRNS model. For statistical specifications, we found a more evident pattern between the two estimation methods, i.e. we observed that the predictive performance of our Bayesian method slowly increased towards the long-end of the yield curve. Concerning the third research question, we found that the statistical equivalence between both Nelson-Siegel bases prevailed in our out-of-sample exercise. Besides, we found additional forecast ability on shorter-term and mid-term maturities when we considered our macro-finance specifications from the last research question.

7 Discussion

Going forward, we present various ideas for future research in this Section. A first extension arises with the implementation of the correlated factor (MF-)AFRNS models. We derived the required yield-adjustments for those specifications, yet left the implementation for further research as our algebraic Bayesian approach to construct the continuous-time parameters from Duffie & Kan (1996) is limited to independent factors. A second extension emerges from the gap in predictive performance between the AFRNS model and the AFNS model. Although the Bayesian estimation of the AFNS model shows improvement over the classical method in the conducted out-of-sample exercise, it would be interesting to see if similar achievements can be obtained via different priors for the AFRNS model. Another idea for future research lies in the statistical limitations that arise from the imposition of no-arbitrage. That is, the imposed freedom of arbitrage prohibits the implementation of a time-varying covariance matrix (Hautsch & Yang (2012)), a time-varying rate of decay (Koop & Korobilis (2010)) and a time-varying unconditional mean (Kozicki & Tinsley (2001)). Comparison between the imposition of arbitrage-free conditions and the implementation of time-varying dynamics within the RNS framework provides a greater insight into the opportunity costs of applying no-arbitrage restrictions.

References

Al-Mohy, A. H., & Higham, N. J. (2010). A new scaling and squaring algorithm for the matrix exponential. *SIAM Journal on Matrix Analysis and Applications*, *31*(3), 970–989.

Altavilla, C., Giacomini, R., & Ragusa, G. (2017). Anchoring the yield curve using survey expectations. *Journal of Applied Econometrics*, *32*(6), 1055–1068.

Altavilla, C., Giannone, D., & Modugno, M. (2017). Low frequency effects of macroeconomic news on government bond yields. *Journal of Monetary Economics*, *92*, 31–46.

Andreasen, M. M., & Christensen, B. J. (2015). The sr approach: A new estimation procedure for non-linear and non-gaussian dynamic term structure models. *Journal of Econometrics*, *184*(2), 420–451.

Andreasen, M. M., & Meldrum, A. (2013). Likelihood inference in non-linear term structure models: the importance of the lower bound. *Available at SSRN 2370449*.

Ang, A., Dong, S., & Piazzesi, M. (2007). *No-arbitrage taylor rules* (Tech. Rep.). National Bureau of Economic Research.

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, *50*(4), 745–787.

Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about gdp growth? *Journal of econometrics*, *131*(1-2), 359–403.

Bauer, M. D., Hamilton, J. D., et al. (2016). Do macro variables help forecast interest rates? *FRBSF Economic Letter*, *20*, 1–5.

Bauer, M. D., & Rudebusch, G. D. (2013). The signaling channel for federal reserve bond purchases. *International Journal of Central Banking*.

Bauer, M. D., Rudebusch, G. D., & Wu, J. C. (2014). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset: Comment. *American Economic Review*, *104*(1), 323–37.

Bech, M. L., & Lengwiler, Y. (2012). The financial crisis and the changing dynamics of the yield curve. *BIS Paper*(65o).

Bernanke, B. S. (2015). The taylor rule: A benchmark for monetary policy? *Ben Bernanke's Blog*, *28*.

Björk, T., & Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, *9*(4), 323–348.

Black, F. (1995). Interest rates as options. *the Journal of Finance*, *50*(5), 1371–1376.

Blinder, A. S. (2000). Monetary policy at the zero lower bound: balancing the risks. *Journal of Money, Credit and Banking*, *32*(4), 1093–1099.

Blinder, A. S., et al. (2010). Quantitative easing: entrance and exit strategies. *Federal Reserve Bank of St. Louis Review*, *92*(6), 465–479.

Bonis, B., Ihrig, J. E., & Wei, M. (2017). The effect of the federal reserve's securities holdings on longer-term interest rates. *Available at SSRN 3300780*.

Çakmaklı, C. (2020). Modeling the density of us yield curve using bayesian semiparametric dynamic nelson-siegel model. *Econometric Reviews*, *39*(1), 71–91.

Caldeira, J. F., Laurini, M. P., & Portugal, M. S. (2010). Bayesian inference applied to dynamic nelson-siegel model with stochastic volatility. *Brazilian Review of Econometrics*, *30*(1), 123–161.

Caldeira, J. F., Moura, G. V., & Santos, A. A. (2016). Predicting the yield curve using forecast combinations. *Computational Statistics & Data Analysis*, *100*, 79–98.

Carter, C. K., & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Charemza, W. W., Hristova*, D., & Burridge, P. (2005). Is inflation stationary? *Applied Economics*, *37*(8), 901–903.

Chib, S., & Ergashev, B. (2009). Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association*, *104*(488), 1324–1337.

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, *49*(4), 327–335.

Christensen, J. H., Diebold, F. X., & Rudebusch. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, *164*(1), 4–20.

Christensen, J. H., & Rudebusch, G. D. (2015). Estimating shadow-rate term structure models with near-zero yields. *Journal of Financial Econometrics*, *13*(2), 226–259.

Christensen, J. H., & Rudebusch, G. D. (2016). *Modeling yields at the zero lower bound: Are shadow rates the solution?* Emerald Group Publishing Limited.

Cieslak, A. (2018). Short-rate expectations and unexpected returns in treasury bonds. *The Review of Financial Studies*, *31*(9), 3265–3306.

Clark, T. E., et al. (2001). Comparing measures of core inflation. *Economic Review-Federal Reserve Bank of Kansas City*, *86*(2), 5–32.

Coroneo, L., Nyholm, K., & Vidova-Koleva, R. (2011). How arbitrage-free is the nelson–siegel model? *Journal of Empirical Finance*, *18*(3), 393–407.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. In *Theory of valuation* (pp. 129–164). World Scientific.

Curves, B. Z.-C. Y. (2005). Technical documentation. *BIS papers*, *25*.

Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The journal of finance*, *55*(5), 1943–1978.

De Jong, P., & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, *82*(2), 339–350.

De Pooter, M., Ravazzolo, F., & Van Dijk, D. J. (2007). Predicting the term structure of interest rates: Incorporating parameter uncertainty, model uncertainty and macroeconomic information. *Model Uncertainty and Macroeconomic Information (October 25, 2007)*.

De Pooter, M., Ravazzolo, F., & Van Dijk, D. J. (2010). Term structure forecasting using macro factors and forecast combination. *FRB International Finance Discussion Paper*(993).

De Rezende, R. B., & Ferreira, M. S. (2013). Modeling and forecasting the yield curve by an extended nelson-siegel class of models: A quantile autoregression approach. *Journal of Forecasting*, *32*(2), 111–123.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, *130*(2), 337–364.

Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic nelson–siegel approach. *Journal of Econometrics*, *146*(2), 351–363.

Diebold, F. X., Rudebusch, & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, *131*(1-2), 309–338.

Diebold, F. X., & Rudebusch, G. D. (2013). *Yield curve modeling and forecasting: the dynamic nelson-siegel approach*. Princeton University Press.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, *57*(1), 405–443.

Duffee, G. R. (2011). Information in (and not in) the term structure. *The Review of Financial Studies*, *24*(9), 2895–2934.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, *6*(4), 379–406.

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–616.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

d'Amico, S., English, W., López-Salido, D., & Nelson, E. (2012). The federal reserve's large-scale asset purchase programmes: rationale and effects. *The Economic Journal*, *122*(564), F415–F446.

Estrella, A., & Mishkin, F. S. (1996). The yield curve as a predictor of us recessions. *Current issues in economics and finance*, *2*(7).

Estrella, A., & Trubin, M. (2006). The yield curve as a leading indicator: Some practical issues. *Current issues in Economics and Finance*, *12*(5).

Feroli, M., Greenlaw, D., Hooper, P., Mishkin, F. S., & Sufi, A. (2017). Language after liftoff: Fed communication away from the zero lower bound. *Research in Economics*, *71*(3), 452–490.

Filipović, D. (1999). A note on the nelson–siegel family. *Mathematical finance*, *9*(4), 349–359.

Fisher, M., & Gilles, C. (1996). Estimating exponential-affine models of the term structure. *Unpublished working paper. Federal Reserve Bank of Atlanta*.

Frankel, J. A., & Lown, C. S. (1994). An indicator of future inflation extracted from the steepness of the interest rate yield curve along its entire length. *The Quarterly Journal of Economics*, *109*(2), 517–530.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, *15*(2), 183–202.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*(6), 721–741.

Geweke, J., Bernardo, J. M., Berger, J. O., Dawid, A. P., & Smith, A. (1992). Bayesian statistics. *Bayesian statistics*.

Greenwood, R., & Vayanos, D. (2014). Bond supply and excess bond returns. *The Review of Financial Studies*, *27*(3), 663–713.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, *54*(8), 2291–2304.

Hamilton, J. D., et al. (1993). *State-space models*. University of California, San Diego. Department of Economics.

Hamilton, J. D., & Wu, J. C. (2012). Identification and estimation of gaussian affine term structure models. *Journal of Econometrics*, *168*(2), 315–331.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Hautsch, N., & Ou, Y. (2008). *Yield curve factors, term structure volatility, and bond risk premia* (Tech. Rep.). SFB 649 discussion paper.

Hautsch, N., & Yang, F. (2012). Bayesian inference in a stochastic volatility nelson–siegel model. *Computational Statistics & Data Analysis*, *56*(11), 3774–3792.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77–105.

Holland, T. E. (1969). Operation twist'and the movement of interest rates and related economic time series. *International Economic Review*, *10*(3), 260–265.

Ihrig, J. E., Klee, E., Li, C., Schulte, B., & Wei, M. (2012). Expectations about the federal reserve's balance sheet and the term structure of interest rates. *Available at SSRN 2193992*.

Kim, Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, *65*(3), 361–393.

Kim, & Wright, J. H. (2005). An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. *Available at SSRN 813267*.

Koop, G., & Korobilis, D. (2010). *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, *74*(1), 119–147.

Koopman, S. J., Mallee, M. I., & Van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, *28*(3), 329–343.

Kozicki, S., & Tinsley, P. A. (2001). Shifting endpoints in the term structure of interest rates. *Journal of monetary Economics*, *47*(3), 613–652.

Krippner, L. (2015). A theoretical foundation for the nelson–siegel class of yield curve models. *Journal of Applied Econometrics*, *30*(1), 97–118.

Krishnamurty, A., & Vissing-Jorgensen, A. (2011). The effects of quantitative easing on interest rates. *Brookings Papers on Economic Activity*, *43*(2).

Laubach, T. (2009). New evidence on the interest rate effects of budget deficits and debt. *Journal of the European Economic Association*, *7*(4), 858–885.

Laurini, M. P., & Caldeira, J. F. (2016). A macro-finance term structure model with multivariate stochastic volatility. *International Review of Economics & Finance*, *44*, 68–90.

Li, C., & Wei, M. (2012). Term structure modelling with supply factors and the federal reserve's large scale asset purchase programs. *Available at SSRN 2191189*.

Litterman. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Litterman. (1991). Common factors affecting bond returns. *Journal of fixed income*, *1*(1), 54–61.

Mendez-Carbajo, D. (2019). Should we fear the inverted yield curve? *Page One Economics®*.

Modigliani, F., & Sutch, R. (1966). Innovations in interest rate policy. *The American Economic Review*, *56*(1/2), 178–197.

Mönch, E. (2012). Term structure surprises: the predictive content of curvature, level, and slope. *Journal of Applied Econometrics*, *27*(4), 574–602.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, 473–489.

Nyholm, K. (2015). A rotated dynamic nelson-siegel model with macro-financial applications. *Available at SSRN 2666978*.

Nyholm, K. (2018). A rotated dynamic nelson-siegel model. *Economic Notes: Review of Banking, Finance and Monetary Economics*, *47*(1), 113–124.

Nyholm, K., & Vidova-Koleva, R. (2012). Nelson–siegel, affine and quadratic yield curve specifications: which one is better at forecasting? *Journal of Forecasting*, *31*(6), 540–564.

Orphanides, A., & Wei, M. (2012). Evolving macroeconomic perceptions and the term structure of interest rates. *Journal of Economic Dynamics and Control*, *36*(2), 239–254.

Pericoli, M., & Taboga, M. (2018). Nearly exact bayesian estimation of non-linear no-arbitrage term structure models. *Bank of Italy Temi di Discussione (Working Paper) No*, *1189*.

Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, *58*(1), 17–29.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–1957 1. *economica*, *25*(100), 283–299.

Rinehart, R. (1955). The equivalence of definitions of a matric function. *The American Mathematical Monthly*, *62*(6), 395–414.

Ross, M. H. (1966). " operation twist": A mistaken policy? *Journal of Political Economy*, *74*(2), 195–199.

Rudebusch, et al. (2018). A review of the fed's unconventional monetary policy. *FRBSF Economic Letter*, *27*.

Rudebusch, & Svensson, L. E. (1999). Policy rules for inflation targeting. In *Monetary policy rules* (pp. 203–262). University of Chicago Press.

Rudebusch, & Wu. (2008). A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal*, *118*(530), 906–926.

Singer, S., & Nelder, J. (2009). Nelder-mead algorithm. *Scholarpedia*, *4*(7), 2928.

Solow, R., Tobin, J., Tobin, J., & Weidenbaum, M. (1987). The kennedy economic reports: Introduction. *Two Revolutions in Economic Policy: The First Economic Reports of Presidents Kennedy and Reagan*, 3–16.

Steeley, J. M. (2014). Forecasting the term structure when short-term rates are near zero. *Journal of Forecasting*, *33*(5), 350–363.

Svensson, L. E. (1994). *Estimating and interpreting forward interest rates: Sweden 1992-1994* (Tech. Rep.). National bureau of economic research.

Svensson, L. E. (1995). Estimating forward interest rates with the extended nelson & siegel method. *Sveriges Riksbank Quarterly Review*, *3*(1), 13–26.

Swanson, E. T., Reichlin, L., & Wright, J. H. (2011). Let's twist again: A high-frequency event-study analysis of operation twist and its implications for qe2 [with comments and discussion]. *Brookings Papers on Economic Activity*, 151–207.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, *82*(398), 528–540.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-rochester conference series on public policy* (Vol. 39, pp. 195–214).

Trolle, A. B., & Schwartz, E. S. (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *The Review of Financial Studies*, *22*(5), 2007–2057.

Trück, S., & Wellmann, D. (2016). Forecasting the term structure of interest rates near the zero bound-a new era? *Available at SSRN 2648164*.

Ullah, W. (2016). Affine term structure model with macroeconomic factors: Do no-arbitrage restriction and macroeconomic factors imply better out-of-sample forecasts? *Journal of Forecasting*, *35*(4), 329–346.

Van Dijk, D., Koopman, S. J., Van der Wel, M., & Wright, J. H. (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, *29*(5), 693–712.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, *5*(2), 177–188.

Williams, D. (1991). *Probability with martingales*. Cambridge university press.

Woodford, M. (2001). The taylor rule and optimal monetary policy. *American Economic Review*, *91*(2), 232–237.

Wright, J. H. (2006). The yield curve and predicting recessions. *Available at SSRN 899538*.

Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review*, *101*(4), 1514–34.

Yu, W.-C., & Zivot, E. (2011). Forecasting the term structures of treasury and corporate yields using dynamic nelson-siegel models. *International Journal of Forecasting*, *27*(2), 579–591.

A Derivation of AFRNS

This part of the appendix provides greater comprehension on the unification of the RNS model with the arbitrage-free framework from Duffie & Kan (1996). In Appendix A.1 we derive the risk-neutral state dynamics of the AFRNS model for which we provide a formal proof in Appendix A.2 as we inversely show the suggested solutions to comply with the Ordinary Differential Equations (ODEs) from Duffie & Kan (1996). Appendix A.3 concludes with an algebraic expression for the yield-adjustment of the AFRNS specification, which linearly relates to the yield-adjustment term of the AFNS model from Christensen et al. (2011).

A.1 Proposition AFRNS under $\mathbb{Q}$-measure

Following the notation from Duffie & Kan (1996), we assume $\boldsymbol{X_t}$ (3×1) to be a Gaussian Ornstein-Uhlenbeck process under the $\mathbb{Q}$-measure with Stochastic Differential Equation (SDE)

$$\mathrm{d}\boldsymbol{X_t} = \boldsymbol{K_t^{\mathbb{Q}}}[\boldsymbol{\theta_t^{\mathbb{Q}}} - \boldsymbol{X}_t]\mathrm{d}t + \boldsymbol{\Sigma_t D_t(X_t)}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}}, \tag{40}$$

where $\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}}$ (3×1) is a multivariate standard Brownian motion under the $\mathbb{Q}$-measure whose filtration is compliant with Williams (1991). We assume a bounded continuous time-varying drift and state dynamics matrix in $\boldsymbol{\theta_t^{\mathbb{Q}}}$ (3×1) and $\boldsymbol{K_t^{\mathbb{Q}}}$ (3×3) respectively. The process' volatility embodies two bounded continuous components, i.e. risk-neutral volatility matrix $\boldsymbol{\Sigma_t}$ (3×3) and diagonal matrix $\boldsymbol{D_t(X_t)}$ (3×3) with elements $\sqrt{\gamma_t^j + (\boldsymbol{\delta_t^j})'\boldsymbol{X_t}}$, where scalar γ_t^j and vector $\boldsymbol{\delta_t^j}$ (3×1) are bounded continuous functions for $j = 1, 2, 3$. Following Christensen et al. (2011) we suppress the possible time-dependence on the parameters from Equation (55) and assume the risk-neutral evolution on the latent factors to be homoskedastic, i.e. $\boldsymbol{D_t(X_t)} = \boldsymbol{I_3}$ $\forall t$.

Duffie & Kan (1996) construct the risk-free rate as an affine function of the underlying latent factors such that

$$r_t = \rho_0 + \boldsymbol{\rho_1'}\boldsymbol{X_t}, \tag{41}$$

where the risk-neutral evolution of the theoretical short rate for the RNS model fits the framework via $\rho_0 = 0$ and $\boldsymbol{\rho_1} = [1\ 0\ 0]'$. Additionally, Duffie & Kan (1996) show zero-coupon bond prices to resolve to exponential functions of the latent states under the $\mathbb{Q}$-measure, i.e.

$$P(t,T) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{\left(-\int_t^T r_u \mathrm{d}u\right)}\right] = e^{A(t,T) + \boldsymbol{B(t,T)}'\boldsymbol{X_t}}, \tag{42}$$

where $A(t,T)$ and $\boldsymbol{B(t,T)}$ refer to time-invariant solutions to the following system of Riccati ODEs

$$\frac{\mathrm{d}A(t,T)}{\mathrm{d}t} = \rho_0 - \boldsymbol{B(t,T)}'(\boldsymbol{K^{\mathbb{Q}}})\boldsymbol{\theta^{\mathbb{Q}}} - \frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{\Sigma}'\boldsymbol{B(t,T)}\boldsymbol{B(t,T)}'\boldsymbol{\Sigma}\right)_{j,j}, \tag{43}$$

$$\frac{\mathrm{d}\boldsymbol{B(t,T)}}{\mathrm{d}t} = \boldsymbol{\rho_1} + (\boldsymbol{K^{\mathbb{Q}}})'\boldsymbol{B(t,T)}, \tag{44}$$

with boundary conditions $A(T,T) = 0$ and $\boldsymbol{B(T,T)} = \mathbf{0}$. Equation (42) implies the underlying yields to be defined as

$$y(t,T) = -\frac{\log P(t,T)}{T-t} = -\frac{A(t,T)}{T-t} - \frac{\boldsymbol{B(t,T)'}}{T-t}\boldsymbol{X_t}, \tag{45}$$

which roughly resembles the structure from Equation (6). Following Christensen et al. (2011), a closest match of the three-factor affine structure from Duffie & Kan (1996) and the factor rotation from Nyholm (2018) yields

$$y(t,T) = X_t^1 + \left[1 - \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}\right]X_t^2 + \left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}\right]X_t^3 - \frac{A(t,T)}{T-t}, \tag{46}$$

where we obtain explicit solutions to the Riccati ODEs via

$$-\frac{B^1(t,T)}{T-t} = 1, \qquad \Leftrightarrow \quad B^1(t,T) = -(T-t), \tag{47}$$

$$-\frac{B^2(t,T)}{T-t} = 1 - \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} \qquad \Leftrightarrow \quad B^2(t,T) = -(T-t) + \frac{1-e^{-\lambda(T-t)}}{\lambda}, \tag{48}$$

$$-\frac{B^3(t,T)}{T-t} = \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \qquad \Leftrightarrow \quad B^3(t,T) = (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda}. \tag{49}$$

We obtain a nearly-exact fit, which is distorted by the time-invariant yield-adjustment. Its algebraic expression is given by

$$A(t,T) = \sum_{j=2}^{3}\left[(\boldsymbol{K}^{\mathbb{Q}}\boldsymbol{\theta}^{\mathbb{Q}})_j \int_t^T B^j(s,T)\mathrm{d}s\right] + \frac{1}{2}\sum_{j=1}^{3}\int_t^T \left(\boldsymbol{\Sigma}'\boldsymbol{B(s,T)B(s,T)'}\boldsymbol{\Sigma}\right)_{j,j}\mathrm{d}s. \tag{50}$$

We refer to Appendix A.3 for explicit completion of Equation (50). In the remainder of this appendix, we explicitly compute the risk-neutral state dynamics from Equation (40). Accordingly, we compute the left-hand-side of Equation (44) via the partial derivatives of

Equations (47)-(49) with respect to t. Hence, we derive the rows of $(\boldsymbol{K}^{\mathbb{Q}})'$ as

$$\frac{\mathrm{d}B^1(t,T)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left[-(T-t)\right] = 1, \tag{51}$$

$$\begin{aligned}\frac{\mathrm{d}B^2(t,T)}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[-(T-t)+\frac{1-e^{-\lambda(T-t)}}{\lambda}\right] = 1-\frac{\lambda e^{-\lambda(T-t)}}{\lambda}\\ &= 1+\lambda\left[\frac{-e^{-\lambda(T-t)}-1+1}{\lambda}\right] = \lambda\left[\frac{1-e^{-\lambda(T-t)}}{\lambda}+(T-t)-(T-t)\right]\\ &= \lambda\left[\frac{1-e^{-\lambda(T-t)}}{\lambda}-(T-t)\right]+-\lambda[-(T-t)] = -\lambda B^1(t,T)+\lambda B^2(t,T),\end{aligned} \tag{52}$$

$$\begin{aligned}\frac{\mathrm{d}B^3(t,T)}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[(T-t)e^{-\lambda(T-t)}-\frac{1-e^{-\lambda(T-t)}}{\lambda}\right] = \lambda(T-t)e^{-\lambda(T-t)}\\ &= \lambda\left[(T-t)e^{-\lambda(T-t)}-\frac{1-e^{-\lambda(T-t)}}{\lambda}+\frac{1-e^{-\lambda(T-t)}}{\lambda}\right]\\ &= \lambda\left[\frac{1-e^{-\lambda(T-t)}}{\lambda}+(T-t)-(T-t)\right]+\lambda B^3(t,T)\\ &= -\lambda B^1(t,T)+\lambda B^2(t,T)+\lambda B^3(t,T).\end{aligned} \tag{53}$$

Hence, explicit filling of Equation (44) yields

$$\frac{\mathrm{d}\boldsymbol{B(t,T)}}{\mathrm{d}t} = \begin{pmatrix}1\\0\\0\end{pmatrix} + \begin{bmatrix}0 & 0 & 0\\ -\lambda & \lambda & 0\\ -\lambda & \lambda & \lambda\end{bmatrix}\begin{pmatrix}B^1(t,T)\\B^2(t,T)\\B^3(t,T)\end{pmatrix}. \tag{54}$$

Additionally, the insertion of $\boldsymbol{K}^{\mathbb{Q}}$ in Equation (55) demonstrates $\boldsymbol{X_t}$ to be Markovian with

$$\begin{pmatrix}\mathrm{d}X_t^1\\ \mathrm{d}X_t^2\\ \mathrm{d}X_t^3\end{pmatrix} = \begin{bmatrix}0 & -\lambda & -\lambda\\ 0 & \lambda & \lambda\\ 0 & 0 & \lambda\end{bmatrix}\left[\begin{pmatrix}\theta_1^{\mathbb{Q}}\\ \theta_2^{\mathbb{Q}}\\ \theta_3^{\mathbb{Q}}\end{pmatrix} - \begin{pmatrix}X_t^1\\ X_t^2\\ X_t^3\end{pmatrix}\right]\mathrm{d}t + \boldsymbol{\Sigma}\begin{pmatrix}\mathrm{d}W_{1,t}^{\mathbb{Q}}\\ \mathrm{d}W_{2,t}^{\mathbb{Q}}\\ \mathrm{d}W_{3,t}^{\mathbb{Q}}\end{pmatrix}, \quad \lambda > 0. \tag{55}$$

A.2 Proof

In this part of the appendix we assume the system of ODEs from Equation (54) for which we inversely show the suggested solutions from Equation (47)-(49) to emerge. Because

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)}\boldsymbol{B(t,T)}\right] = e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)}\frac{\mathrm{d}\boldsymbol{B(t,T)}}{\mathrm{d}t} - (\boldsymbol{K}^{\mathbb{Q}})'e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)}\boldsymbol{B(t,T)}, \tag{56}$$

it follows from Equation (44) that

$$\int_t^T \frac{\mathrm{d}}{\mathrm{d}s}\left[e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-s)}\boldsymbol{B(s,T)}\right]\mathrm{d}s = \int_t^T e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-s)}\boldsymbol{\rho_1}\mathrm{d}s, \tag{57}$$

which further simplifies to

$$\boldsymbol{B(t,T)} = e^{-(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} \int_t^T e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-s)} \boldsymbol{\rho_1} \mathrm{d}s \tag{58}$$

using boundary conditions $A(T,T) = 0$ and $\boldsymbol{B(T,T)} = \boldsymbol{0}$. We construct both $(\boldsymbol{K}^{\mathbb{Q}})'$ and $\boldsymbol{\rho_1}$ via our proposition on the risk-neutral state dynamics on the AFRNS model from Appendix A.1 as

$$(\boldsymbol{K}^{\mathbb{Q}})' = \begin{bmatrix} 0 & 0 & 0 \\ -\lambda & \lambda & 0 \\ -\lambda & \lambda & \lambda \end{bmatrix}, \qquad \boldsymbol{\rho_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{59}$$

where explicit solutions for the matrix exponentials $e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)}$ and $e^{-(\boldsymbol{K}^{\mathbb{Q}})'(T-t)}$ reside. Hence, we resort to the non-diagonal generalization of Sylvester's formula (e.g. see Rinehart (1955) for an explanation in greater detail). Eigenvector decomposition of $(\boldsymbol{K}^{\mathbb{Q}})'$ yields eigenvalues $q_1 = 0$ and $q_2 = \lambda$, such that linearization of $e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)}$ settles to

$$\begin{aligned} e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} &= \boldsymbol{B_{1,1}} e^{q_1(T-t)} + \boldsymbol{B_{1,2}}(T-t)e^{q_1 t} + \boldsymbol{B_{2,1}} e^{q_2(T-t)} + \boldsymbol{B_{2,2}}(T-t)e^{q_2(T-t)} \\ &= \boldsymbol{B_{1,1}} + \boldsymbol{B_{1,2}}(T-t) + \boldsymbol{B_{2,1}} e^{\lambda(T-t)} + \boldsymbol{B_{2,2}}(T-t)e^{\lambda(T-t)}, \end{aligned} \tag{60}$$

where closed-form solutions for the coefficient matrices $\{\boldsymbol{B_{i,j}}\}_{i,j=1,2}$ solve the linear system. Accordingly, we identify the system via the first three differentials of Equation (60) with respect to $T-t$, i.e.

$$[(\boldsymbol{K}^{\mathbb{Q}})']e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} = \boldsymbol{B_{1,2}} + \lambda \boldsymbol{B_{2,1}} e^{\lambda(T-t)} + (\lambda(T-t)+1)\boldsymbol{B_{2,2}} e^{\lambda(T-t)}, \tag{61}$$

$$[(\boldsymbol{K}^{\mathbb{Q}})']^2 e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} = \lambda^2 \boldsymbol{B_{2,1}} e^{\lambda(T-t)} + \lambda(2+\lambda(T-t))\boldsymbol{B_{2,2}} e^{\lambda(T-t)}, \tag{62}$$

$$[(\boldsymbol{K}^{\mathbb{Q}})']^3 e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} = \lambda^3 \boldsymbol{B_{2,1}} e^{\lambda(T-t)} + \lambda^2(3+\lambda(T-t))\boldsymbol{B_{2,2}} e^{\lambda(T-t)}, \tag{63}$$

where inserting $T = t$ provides analytical solutions for the coefficient matrices $\{\boldsymbol{B_{i,j}}\}_{i,j=1,2}$ as it simplifies Equations (60)-(63) to

$$\boldsymbol{I}_3 = \boldsymbol{B_{1,1}} + \boldsymbol{B_{2,1}} \quad \Leftrightarrow \quad \boldsymbol{B_{1,1}} = \boldsymbol{I}_3 - \boldsymbol{B_{2,1}}, \tag{64}$$

$$\boldsymbol{\Omega} = \boldsymbol{B_{1,2}} + \lambda \boldsymbol{B_{2,1}} + \boldsymbol{B_{2,2}} \quad \Leftrightarrow \quad \boldsymbol{B_{1,2}} = \boldsymbol{0}, \tag{65}$$

$$\boldsymbol{\Omega}^2 = \lambda^2 \boldsymbol{B_{2,1}} + 2\lambda \boldsymbol{B_{2,2}} \quad \Leftrightarrow \quad \boldsymbol{B_{2,1}} = \frac{1}{\lambda^2}(\boldsymbol{\Omega}^2 - 2\lambda \boldsymbol{B_{2,2}}), \tag{66}$$

$$\boldsymbol{\Omega}^3 = \lambda^3 \boldsymbol{B_{2,1}} + 3\lambda^2 \boldsymbol{B_{2,2}} \quad \Leftrightarrow \quad \boldsymbol{B_{2,2}} = \frac{1}{\lambda^2}(\boldsymbol{\Omega}^3 - \lambda \boldsymbol{\Omega}^2). \tag{67}$$

Thus, from Equation (60) we obtain algebraic expressions for the matrix exponentials via

$$e^{(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} = \begin{bmatrix} 1 & 0 & 0 \\ 1 - e^{\lambda(T-t)} & e^{\lambda(T-t)} & 0 \\ -\lambda(T-t)e^{\lambda(T-t)} & \lambda(T-t)e^{\lambda(T-t)} & e^{\lambda(T-t)} \end{bmatrix} \tag{68}$$

and

$$e^{-(\boldsymbol{K}^{\mathbb{Q}})'(T-t)} = \begin{bmatrix} 1 & 0 & 0 \\ 1-e^{-\lambda(T-t)} & e^{-\lambda(T-t)} & 0 \\ \lambda(T-t)e^{-\lambda(T-t)} & -\lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)}. \end{bmatrix}. \tag{69}$$

Hence, all components of Equation (58) are known. Explicit filling yields

$$\boldsymbol{B}(\boldsymbol{t},\boldsymbol{T}) = -\begin{bmatrix} 1 & 0 & 0 \\ 1-e^{-\lambda(T-t)} & e^{-\lambda(T-t)} & 0 \\ \lambda(T-t)e^{-\lambda(T-t)} & -\lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)}, \end{bmatrix} \int_t^T \begin{pmatrix} 1 \\ 1-e^{\lambda(T-s)} \\ -\lambda(T-t)e^{\lambda(T-s)} \end{pmatrix} \mathrm{d}s \tag{70}$$

where the integrals simplify to

$$\int_t^T \mathrm{d}s = T-t, \tag{71}$$

$$\int_t^T 1-e^{\lambda(T-s)}\mathrm{d}s = T-t+\frac{1-e^{\lambda(T-t)}}{\lambda}, \tag{72}$$

$$\int_t^T -\lambda(T-t)e^{\lambda(T-s)}\mathrm{d}s = -(T-t)e^{\lambda(T-t)} - \frac{1-e^{\lambda(T-t)}}{\lambda}. \tag{73}$$

Thus, a final expression for Equation (58) resolves to

$$\begin{aligned} \boldsymbol{B}(\boldsymbol{t},\boldsymbol{T}) &= -\begin{bmatrix} 1 & 0 & 0 \\ 1-e^{-\lambda(T-t)} & e^{-\lambda(T-t)} & 0 \\ \lambda(T-t)e^{-\lambda(T-t)} & -\lambda(T-t)e^{-\lambda(T-t)} & e^{-\lambda(T-t)} \end{bmatrix} \begin{pmatrix} T-t \\ T-t+\frac{1-e^{\lambda(T-t)}}{\lambda} \\ -(T-t)e^{\lambda(T-t)} - \frac{1-e^{\lambda(T-t)}}{\lambda} \end{pmatrix} \\ &= \begin{pmatrix} -(T-t) \\ -(T-t)+\frac{1-e^{-\lambda(T-t)}}{\lambda} \\ (T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda} \end{pmatrix}, \end{aligned} \tag{74}$$

which coincides with the proposed solutions to the system of ODEs from Equations (47)-(49). Hence, the proposition on the risk-neutral state dynamics of the AFRNS model from Appendix A.1 is proved.

A.3 Yield-Adjustment Term under the $\mathbb{Q}$-measure

We identify the AFRNS model by restraining its risk-neutral drift to be nonexistent, i.e. $\boldsymbol{\theta}^{\mathbb{Q}} = \mathbf{0}$, which further simplifies Equation (50) to

$$\begin{aligned}
\frac{A(t,T)}{T-t} &= \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^{3}\Big(\boldsymbol{\Sigma}'\boldsymbol{B}(\boldsymbol{s},\boldsymbol{T})\boldsymbol{B}(\boldsymbol{s},\boldsymbol{T})'\boldsymbol{\Sigma}\Big)_{j,j}\mathrm{d}s \\
&= \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^{3}\left(\begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix}\begin{pmatrix}B^1(s,T)\\ B^2(s,T)\\ B^3(s,T)\end{pmatrix}\Big(B^1(s,T) \quad B^2(s,T) \quad B^3(s,T)\Big)\right. \\
&\qquad \left.\begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix}\right)_{j,j}\mathrm{d}s \\
&= \frac{\bar{A}}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2\mathrm{d}s + \frac{\bar{B}}{2}\frac{1}{T-t}\int_t^T B^2(s,T)^2\mathrm{d}s + \frac{\bar{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)^2\mathrm{d}s \\
&\quad + \bar{D}\frac{1}{T-t}\int_t^T B^1(s,T)B^2(s,T)\mathrm{d}s + \bar{E}\frac{1}{T-t}\int_t^T B^1(s,T)B^3(s,T)\mathrm{d}s \\
&\quad + \bar{F}\frac{1}{T-t}\int_t^T B^2(s,T)B^3(s,T)\mathrm{d}s. \qquad (75)
\end{aligned}$$

The alphabetic coefficients are functions of the risk-neutral volatilities such that

$$\bar{A} = \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2, \qquad (76)$$

$$\bar{B} = \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2, \qquad (77)$$

$$\bar{C} = \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2, \qquad (78)$$

$$\bar{D} = \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23}, \qquad (79)$$

$$\bar{E} = \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33}, \qquad (80)$$

$$\bar{F} = \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33}. \qquad (81)$$

Christensen et al. (2011) accordingly, we obtain a closed-form expression for the yield-adjustment term of the AFRNS model after solving the six smaller integrals. Hence, the first is

$$\begin{aligned}
I_1^{RNS} &= \frac{\bar{A}}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2\mathrm{d}s \\
&= \frac{\bar{A}}{2}\frac{1}{T-t}\int_t^T (T-s)^2\mathrm{d}s = \frac{\bar{A}}{6}(T-t)^2. \qquad (82)
\end{aligned}$$

The second is

$$
\begin{aligned}
I_2^{RNS} &= \frac{\bar{B}}{2}\frac{1}{T-t}\int_t^T B^2(s,T)^2\mathrm{d}s \\
&= \frac{\bar{B}}{2}\frac{1}{T-t}\int_t^T \left[-(T-s) + \frac{1-e^{-\lambda(T-s)}}{\lambda}\right]^2 \mathrm{d}s \\
&= \frac{\bar{B}}{2}\frac{1}{T-t}\left[\int_t^T (T-s)^2\mathrm{d}s - 2\int_t^T \Big((T-s)\Big[\frac{1-e^{-\lambda(T-s)}}{\lambda}\Big]\Big)\mathrm{d}s + \int_t^T \Big(\frac{1-e^{-\lambda(T-s)}}{\lambda}\Big)^2\mathrm{d}s\right] \\
&= \frac{\bar{B}}{6}(T-t)^2 - \bar{B}\Big[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big] \\
&\quad + \bar{B}\Big[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big].
\end{aligned} \tag{83}
$$

The third is

$$
\begin{aligned}
I_3^{RNS} &= \frac{\bar{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)^2\mathrm{d}s \\
&= \frac{\bar{C}}{2}\frac{1}{T-t}\int_t^T \left[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right]^2 \mathrm{d}s \\
&= \bar{C}\left[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} - \frac{3}{4\lambda^2}e^{-2\lambda(T-t)}\right. \\
&\quad \left. - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right].
\end{aligned} \tag{84}
$$

The fourth is

$$
\begin{aligned}
I_4^{RNS} &= \frac{\bar{D}}{T-t}\int_t^T B^1(s,T)B^2(s,T)\mathrm{d}s \\
&= \frac{\bar{D}}{T-t}\int_t^T \left[\Big(-(T-s)\Big)\times\Big(-(T-s) + \frac{1-e^{-\lambda(T-s)}}{\lambda}\Big)\right]\mathrm{d}s \\
&= \frac{\bar{D}}{T-t}\left[\int_t^T (T-s)^2\mathrm{d}s - \int_t^T \Big((T-s)\Big[\frac{1-e^{-\lambda(T-s)}}{\lambda}\Big]\Big)\mathrm{d}s\right] \\
&= \frac{\bar{D}}{3}(T-t)^2 - \bar{D}\Big[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\Big].
\end{aligned} \tag{85}
$$

The fifth is

$$
\begin{aligned}
I_5^{RNS} &= \frac{\bar{E}}{T-t}\int_t^T B^1(s,T)B^3(s,T)\mathrm{d}s \\
&= \frac{\bar{E}}{T-t}\int_t^T \left[\Big(-(T-s)\Big)\times\Big((T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda}\Big)\right]\mathrm{d}s \\
&= \bar{E}\left[\frac{3}{\lambda^2}e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right].
\end{aligned} \tag{86}
$$

The sixth is

$$
\begin{aligned}
I_6^{RNS} &= \frac{\bar{F}}{T-t}\int_t^T B^2(s,T)B^3(s,T)\mathrm{d}s \\
&= \frac{\bar{F}}{T-t}\int_t^T \left[\Big(-(T-s)+\frac{1-e^{-\lambda(T-s)}}{\lambda}\Big)\times\Big((T-t)e^{-\lambda(T-t)}-\frac{1-e^{-\lambda(T-t)}}{\lambda}\Big)\right]\mathrm{d}s \\
&= \bar{F}\left[\frac{3}{\lambda^2}e^{-\lambda(T-t)}+\frac{1}{2\lambda}(T-t)+\frac{1}{\lambda}(T-t)e^{-\lambda(T-t)}-\frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right] \\
&\quad -\bar{F}\left[\frac{1}{\lambda^2}+\frac{1}{\lambda^2}e^{-\lambda(T-t)}-\frac{1}{2\lambda^2}e^{-2\lambda(T-t)}-\frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}+\frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right]. \quad (87)
\end{aligned}
$$

An algebraic expression for the yield-adjustment term greatly facilitates empirical implementation. Moreover, the yield-adjustment terms of the AFRNS model and the AFNS model from Christensen et al. (2011) are linearly related. If we define I_i^{DNS} as the i-th identity from Christensen et al. (2011) where $i = 1, 2, \dots 6$, it follows that

$$
\begin{pmatrix} I_1^{RNS} \\ I_2^{RNS} \\ I_3^{RNS} \\ I_4^{RNS} \\ I_5^{RNS} \\ I_6^{RNS} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{\bar{B}}{\bar{A}} & 1 & 0 & -\frac{\bar{B}}{\bar{D}} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{2\bar{D}}{\bar{A}} & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{\bar{F}}{\bar{E}} & -1 \end{bmatrix} \begin{pmatrix} I_1^{DNS} \\ I_2^{DNS} \\ I_3^{DNS} \\ I_4^{DNS} \\ I_5^{DNS} \\ I_6^{DNS} \end{pmatrix}, \quad (88)
$$

where an alternative representation of Equation (75) via the six integrals from Christensen et al. (2011) yields

$$
\frac{A(t,T)}{T-t} = \Big(1+\frac{\bar{B}+2\bar{D}}{\bar{A}}\Big)I_1^{DNS} + I_2^{DNS} + I_3^{DNS} - \Big(1+\frac{\bar{B}}{\bar{D}}\Big)I_4^{DNS} + \Big(1+\frac{\bar{F}}{\bar{E}}\Big)I_5^{DNS} - I_6^{DNS}. \quad (89)
$$

Thus, a linear relation between the factor loadings of the DNS model and the RNS model extends to the yield-adjustment, although the estimated values of the risk-neutral volatilities between the two versions may deviate.

In line with Christensen et al. (2011), not all nine underlying volatility components are identified as we solely face six underlying linear identities in Equations (76)-(81). Hence, the maximally flexible risk-neutral volatility matrix reduces to a lower-triangular one, i.e.

$$
\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}. \quad (90)
$$

B Imposing Restrictions on AFRNS

In this part of the appendix we derive the additional constraints required to impose both variants of the AFRNS model from Section 2.3 to the canonical $A_0(3)$ class of models from Dai & Singleton (2000), i.e. a generic class of no-arbitrage specifications with a three-factor latent state diffusion and zero square-root volatility processes. The extra restrictions arise as we prohibit affine transformations of a three-factor state diffusion to yield equivalent representations of the same fundamental specification.

Thus, we follow Christensen et al. (2011) and define the risk-neutral SDE for an arbitrary three-dimensional latent state vector $\boldsymbol{Y_t}$ as

$$\mathrm{d}\boldsymbol{Y_t} = \boldsymbol{K_Y^{\mathbb{Q}}}[\boldsymbol{\theta_Y^{\mathbb{Q}}} - \boldsymbol{Y_t}]\mathrm{d}t + \boldsymbol{\Sigma_Y}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}}, \tag{91}$$

which fits the framework from Duffie & Kan (1996) via suppressed time-dependence on the parameters. In order to derive the mandatory additional restrictions to guarantee the uniqueness of the risk-neutral process from Equation (91), we consider an affine transformation of the state variables via a nonsingular matrix $\boldsymbol{C}$ (3×3) and a vector of constants $\boldsymbol{\varepsilon}$ (3×1), i.e. $\mathcal{T}_C : \boldsymbol{Z_t} = \boldsymbol{C}\boldsymbol{Y_t} + \boldsymbol{\epsilon}$. From Itô's Lemma it follows that

$$\begin{aligned}
\mathrm{d}\boldsymbol{Z_t} &= \boldsymbol{C}\mathrm{d}\boldsymbol{Y_t} = \boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}}[\boldsymbol{\theta_Y^{\mathbb{Q}}} - \boldsymbol{Y_t}]\mathrm{d}t + \boldsymbol{C}\boldsymbol{\Sigma_Y}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}} = [\boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}}\boldsymbol{\theta_Y^{\mathbb{Q}}} - \boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}}\boldsymbol{Y_t}]\mathrm{d}t + \boldsymbol{C}\boldsymbol{\Sigma_Y}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}} \\
&= \boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}}\boldsymbol{C}^{-1}[\boldsymbol{C}\boldsymbol{\theta_Y^{\mathbb{Q}}} - \boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}} - \boldsymbol{\epsilon} + \boldsymbol{\epsilon}]\mathrm{d}t + \boldsymbol{C}\boldsymbol{\Sigma_Y}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}} \\
&= \boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}}\boldsymbol{C}^{-1}[(\boldsymbol{C}\boldsymbol{\theta_Y^{\mathbb{Q}}} + \boldsymbol{\epsilon}) - \boldsymbol{Z_t}]\mathrm{d}t + \boldsymbol{C}\boldsymbol{\Sigma_Y}\mathrm{d}\boldsymbol{W_t^{\mathbb{Q}}},
\end{aligned} \tag{92}$$

such that $\boldsymbol{Z_t}$ itself follows a three-factor state diffusion under the risk-neutral measure with drift $\boldsymbol{C}\boldsymbol{\theta_Y^{\mathbb{Q}}} + \boldsymbol{\epsilon}$ (3×1), state dynamics $\boldsymbol{C}\boldsymbol{K_Y^{\mathbb{Q}}}\boldsymbol{C}^{-1}$ (3×3) and volatility matrix $\boldsymbol{C}\boldsymbol{\Sigma_Y}$ (3×3). A similar result holds under the $\mathbb{P}$-measure. For the risk-neutral evolution on the short rate it holds that

$$\begin{aligned}
r_t &= \rho_0^Y + (\boldsymbol{\rho_1^Y})'\boldsymbol{Y_t} = \rho_0^Y + (\boldsymbol{\rho_1^Y})\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{Y_t} \\
&= \rho_0^Y + (\boldsymbol{\rho_1^Y})\boldsymbol{C}^{-1}[\boldsymbol{C}\boldsymbol{Y_t} + \boldsymbol{\epsilon} - \boldsymbol{\epsilon}] = \rho_0^Y - (\boldsymbol{\rho_1^Y})'\boldsymbol{C}^{-1}\boldsymbol{\epsilon} + (\boldsymbol{\rho_1^Y})'\boldsymbol{C}^{-1}\boldsymbol{Z_t},
\end{aligned} \tag{93}$$

which relates the short rate processes for both $\boldsymbol{Y_t}$ and $\boldsymbol{Z_t}$ via constant term $\rho_0^Z = \rho_0^Y - (\boldsymbol{\rho_1^Y})\boldsymbol{C}^{-1}\boldsymbol{\epsilon}$ and loading term $(\boldsymbol{\rho_1^Z})' = (\boldsymbol{\rho_1^Y})'\boldsymbol{C}^{-1}$. Both $\boldsymbol{Y_t}$ and $\boldsymbol{Z_t}$ are equivalent compositions of the same underlying model as they are affine latent factor state diffusions with ditto dynamics on the short rate process in Equation (93). Hence, we refer to the transformation $\mathcal{T}_C : \boldsymbol{Z_t} = \boldsymbol{C}\boldsymbol{Y_t} + \boldsymbol{\epsilon}$ as an affine invariant transformation.

In order to identify the system under the $\mathbb{Q}$-measure, Dai & Singleton (2000) restrict the state dynamics $\boldsymbol{K_Y^{\mathbb{Q}}}$ (3×3) to be upper-triangular, the risk-neutral volatility $\boldsymbol{\Sigma_Y}$ (3×3) to be the identity matrix and the unconditional mean $\boldsymbol{\theta_Y^{\mathbb{Q}}}$ (3×1) to be the null vector. Hence, we

obtain $\mathbb{Q}$-dynamics

$$\begin{pmatrix} \mathrm{d}Y_t^1 \\ \mathrm{d}Y_t^2 \\ \mathrm{d}Y_t^3 \end{pmatrix} = - \begin{bmatrix} \kappa_{11}^{Y,\mathbb{Q}} & \kappa_{12}^{Y,\mathbb{Q}} & \kappa_{13}^{Y,\mathbb{Q}} \\ 0 & \kappa_{22}^{Y,\mathbb{Q}} & \kappa_{23}^{Y,\mathbb{Q}} \\ 0 & 0 & \kappa_{33}^{Y,\mathbb{Q}} \end{bmatrix} \begin{pmatrix} Y_t^1 \\ Y_t^2 \\ Y_t^3 \end{pmatrix} \mathrm{d}t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{Q}} \\ \mathrm{d}W_t^{2,\mathbb{Q}} \\ \mathrm{d}W_t^{3,\mathbb{Q}} \end{pmatrix}, \tag{94}$$

and $\mathbb{P}$-dynamics

$$\begin{pmatrix} \mathrm{d}Y_t^1 \\ \mathrm{d}Y_t^2 \\ \mathrm{d}Y_t^3 \end{pmatrix} = \begin{bmatrix} \kappa_{11}^{Y,\mathbb{P}} & \kappa_{12}^{Y,\mathbb{P}} & \kappa_{13}^{Y,\mathbb{P}} \\ \kappa_{21}^{Y,\mathbb{P}} & \kappa_{22}^{Y,\mathbb{P}} & \kappa_{23}^{Y,\mathbb{P}} \\ \kappa_{31}^{Y,\mathbb{P}} & \kappa_{32}^{Y,\mathbb{P}} & \kappa_{33}^{Y,\mathbb{P}} \end{bmatrix} \left[\begin{pmatrix} \theta_1^{Y,\mathbb{P}} \\ \theta_2^{Y,\mathbb{P}} \\ \theta_3^{Y,\mathbb{P}} \end{pmatrix} - \begin{pmatrix} Y_t^1 \\ Y_t^2 \\ Y_t^3 \end{pmatrix} \right] \mathrm{d}t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{P}} \\ \mathrm{d}W_t^{2,\mathbb{P}} \\ \mathrm{d}W_t^{3,\mathbb{P}} \end{pmatrix}, \tag{95}$$

with instantaneous risk-free rate

$$r_t = \rho_0^Y + \rho_{1,1}^Y Y_t^1 + \rho_{1,2}^Y Y_t^2 + \rho_{1,3}^Y Y_t^3. \tag{96}$$

Thus, for the risk-neutral measure we find six free parameters, where we distinguish twelve free parameters for the real-world measure. Additionally, the utmost resilient canonical representation of the $A_0(3)$ class of models offers 22 free parameters as we observe four free parameters on the risk-free rate process in Equation (96).

However, the restrictions in Equations (94)-(95) do not suffice to guarantee the uniqueness of the risk-neutral state diffusion of $\boldsymbol{Y_t}$ under the proposed affine transformations. Hence, we provide additional restrictions to identify the independent factor AFRNS model and the correlated factor AFRNS model in Appendices B.1 and B.2 respectively.

B.1 Independent Factor AFRNS

The independent factor AFRNS model has $\mathbb{P}$-dynamics

$$\begin{pmatrix} \mathrm{d}Z_t^1 \\ \mathrm{d}Z_t^2 \\ \mathrm{d}Z_t^3 \end{pmatrix} = - \begin{bmatrix} \kappa_{11}^{Z,\mathbb{P}} & 0 & 0 \\ 0 & \kappa_{22}^{Z,\mathbb{P}} & 0 \\ 0 & 0 & \kappa_{33}^{Z,\mathbb{P}} \end{bmatrix} \left[\begin{pmatrix} \theta_1^{Z,\mathbb{P}} \\ \theta_2^{Z,\mathbb{P}} \\ \theta_3^{Z,\mathbb{P}} \end{pmatrix} - \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} \right] \mathrm{d}t + \begin{bmatrix} \sigma_{11}^Z & 0 & 0 \\ 0 & \sigma_{22}^Z & 0 \\ 0 & 0 & \sigma_{33}^Z \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{P}} \\ \mathrm{d}W_t^{2,\mathbb{P}} \\ \mathrm{d}W_t^{3,\mathbb{P}} \end{pmatrix}, \tag{97}$$

and $\mathbb{Q}$-dynamics

$$\begin{pmatrix} \mathrm{d}Z_t^1 \\ \mathrm{d}Z_t^2 \\ \mathrm{d}Z_t^3 \end{pmatrix} = - \begin{bmatrix} 0 & -\lambda & -\lambda \\ 0 & \lambda & \lambda \\ 0 & 0 & \lambda \end{bmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} \mathrm{d}t + \begin{bmatrix} \sigma_{11}^Z & 0 & 0 \\ 0 & \sigma_{22}^Z & 0 \\ 0 & 0 & \sigma_{33}^Z \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{Q}} \\ \mathrm{d}W_t^{2,\mathbb{Q}} \\ \mathrm{d}W_t^{3,\mathbb{Q}} \end{pmatrix}, \tag{98}$$

with explicit completion of Equation (96) via $r_t = Z_t^1$. The independent factor AFRNS model offers limited flexibility with ten freely determined parameters and thus requires twelve parameter restrictions on the $A_0(3)$ class of models. One may verify that the affine

transformation $\mathcal{T}_C : \boldsymbol{Z_t} = \boldsymbol{C}\boldsymbol{Y_t} + \boldsymbol{\epsilon}$ with

$$\boldsymbol{C} = \begin{bmatrix} \sigma_{11}^Z & 0 & 0 \\ 0 & \sigma_{22}^Z & 0 \\ 0 & 0 & \sigma_{33}^Z \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{99}$$

yields the independent factor AFRNS model. We derive mean-reversion matrix $\boldsymbol{K}_{\boldsymbol{Y}}^{\mathbb{Q}}$ using Equation (92) as

$$\boldsymbol{K}_{\boldsymbol{Y}}^{\mathbb{Q}} = \begin{bmatrix} \frac{1}{\sigma_{11}^Z} & 0 & 0 \\ 0 & \frac{1}{\sigma_{22}^Z} & 0 \\ 0 & 0 & \frac{1}{\sigma_{33}^Z} \end{bmatrix} \begin{bmatrix} 0 & -\lambda & -\lambda \\ 0 & \lambda & \lambda \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} \sigma_{11}^Z & 0 & 0 \\ 0 & \sigma_{22}^Z & 0 \\ 0 & 0 & \sigma_{33}^Z \end{bmatrix} \tag{100}$$

$$= \begin{bmatrix} 0 & -\lambda\frac{\sigma_{22}^Z}{\sigma_{11}^Z} & -\lambda\frac{\sigma_{33}^Z}{\sigma_{11}^Z} \\ 0 & \lambda & \lambda\frac{\sigma_{33}^Z}{\sigma_{22}^Z} \\ 0 & 0 & \lambda \end{bmatrix}, \tag{101}$$

where the imposition of three extra parameter restrictions is required for $\boldsymbol{K}_{\boldsymbol{Y}}^{\mathbb{Q}}$ to be compliant with the independent factor AFRNS model, i.e.

$$\kappa_{11}^{Y,\mathbb{Q}} = 0, \quad \kappa_{22}^{Y,\mathbb{Q}} = \kappa_{33}^{Y,\mathbb{Q}}, \quad \kappa_{23}^{Y,\mathbb{Q}} = \kappa_{22}^{Y,\mathbb{Q}}\left(\frac{\kappa_{13}^{Y,\mathbb{Q}}}{\kappa_{12}^{Y,\mathbb{Q}}}\right). \tag{102}$$

Hence, we enforce $\kappa_{12}^{Y,\mathbb{Q}}$ and $\kappa_{13}^{Y,\mathbb{Q}}$ to have opposed signs relative to $\kappa_{22}^{Y,\mathbb{Q}}$ and $\kappa_{33}^{Y,\mathbb{Q}}$, although their absolute magnitudes can vary independently. Six more restrictions emerge from $\boldsymbol{K}_{\boldsymbol{Y}}^{\mathbb{P}}$ being diagonal as its derivation settles to a three-factor product of diagonal matrices (e.g. see Equation (92)).

Consequently, we shift focus towards the evolution of the short rate. In all discussed models with the factor rotation from Nyholm (2018) the first latent factor resembles the short rate process. Hence, we impose

$$\rho_0^Z = 0, \qquad \boldsymbol{\rho_1^Z} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{103}$$

in the affine transformation of Equation (93). For the constant term we have $\rho_0^Y = \rho_0^Z = 0$,

where for the loading term $(\boldsymbol{\rho}_{\mathbf{1}}^{\boldsymbol{Z}})' = (\boldsymbol{\rho}_{\mathbf{1}}^{\boldsymbol{Y}})'\boldsymbol{C}^{-1}$ implies that

$$
\begin{aligned}
(\boldsymbol{\rho}_{\mathbf{1}}^{\boldsymbol{Y}})' = (\boldsymbol{\rho}_{\mathbf{1}}^{\boldsymbol{Z}})'\boldsymbol{C} &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} \sigma_{11}^{Z} & 0 & 0 \\ 0 & \sigma_{22}^{Z} & 0 \\ 0 & 0 & \sigma_{33}^{Z} \end{bmatrix} \\
&= \begin{pmatrix} \sigma_{11}^{Z} & 0 & 0 \end{pmatrix}. \qquad (104)
\end{aligned}
$$

Thus, the final three restrictions correspond to $\rho_0^Y = \rho_{1,2}^Y = \rho_{1,3}^Y = 0$.

B.2 Correlated Factor AFRNS

The correlated factor AFRNS model has $\mathbb{P}$-dynamics

$$
\begin{pmatrix} \mathrm{d}Z_t^1 \\ \mathrm{d}Z_t^2 \\ \mathrm{d}Z_t^3 \end{pmatrix} = \begin{bmatrix} \kappa_{11}^{Z,\mathbb{P}} & \kappa_{12}^{Z,\mathbb{P}} & \kappa_{13}^{Z,\mathbb{P}} \\ \kappa_{21}^{Z,\mathbb{P}} & \kappa_{22}^{Z,\mathbb{P}} & \kappa_{23}^{Z,\mathbb{P}} \\ \kappa_{31}^{Z,\mathbb{P}} & \kappa_{32}^{Z,\mathbb{P}} & \kappa_{33}^{Z,\mathbb{P}} \end{bmatrix} \left[\begin{pmatrix} \theta_1^{Z,\mathbb{P}} \\ \theta_2^{Z,\mathbb{P}} \\ \theta_3^{Z,\mathbb{P}} \end{pmatrix} - \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} \right] \mathrm{d}t + \begin{bmatrix} \sigma_{11}^{Z} & \sigma_{12}^{Z} & \sigma_{13}^{Z} \\ 0 & \sigma_{22}^{Z} & \sigma_{23}^{Z} \\ 0 & 0 & \sigma_{33}^{Z} \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{P}} \\ \mathrm{d}W_t^{2,\mathbb{P}} \\ \mathrm{d}W_t^{3,\mathbb{P}} \end{pmatrix}, \quad (105)
$$

and $\mathbb{Q}$-dynamics

$$
\begin{pmatrix} \mathrm{d}Z_t^1 \\ \mathrm{d}Z_t^2 \\ \mathrm{d}Z_t^3 \end{pmatrix} = - \begin{bmatrix} 0 & -\lambda & -\lambda \\ 0 & \lambda & \lambda \\ 0 & 0 & \lambda \end{bmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix} \mathrm{d}t + \begin{bmatrix} \sigma_{11}^{Z} & \sigma_{12}^{Z} & \sigma_{13}^{Z} \\ 0 & \sigma_{22}^{Z} & \sigma_{23}^{Z} \\ 0 & 0 & \sigma_{33}^{Z} \end{bmatrix} \begin{pmatrix} \mathrm{d}W_t^{1,\mathbb{Q}} \\ \mathrm{d}W_t^{2,\mathbb{Q}} \\ \mathrm{d}W_t^{3,\mathbb{Q}} \end{pmatrix}, \quad (106)
$$

which offers greater adaptability compared to the independent factor AFRNS model due to nineteen freely determined parameters. Hence, we need three extra parameter restrictions to identify the system. Following the steps of Appendix B.1, we verify that the affine invariant transformation $\mathcal{T}_C(\boldsymbol{Y_t}) = \boldsymbol{C}\boldsymbol{Y_t} + \boldsymbol{\epsilon}$ with

$$
\boldsymbol{C} = \begin{bmatrix} \sigma_{11}^{Z} & \sigma_{12}^{Z} & \sigma_{13}^{Z} \\ 0 & \sigma_{22}^{Z} & \sigma_{23}^{Z} \\ 0 & 0 & \sigma_{33}^{Z} \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad (107)
$$

converts the canonical representation of the $A_0(3)$ class of models to the correlated factor AFRNS model. Using Equation (92) we derive its risk-neutral state dynamics matrix as

$$\begin{aligned}\boldsymbol{K}_{\boldsymbol{Y}}^{\mathbb{Q}} &= \begin{bmatrix} \frac{1}{\sigma_{11}^Z} & -\frac{\sigma_{12}^Z}{\sigma_{11}^Z\sigma_{22}^Z} & \frac{\sigma_{23}^Z\sigma_{12}^Z-\sigma_{13}^Z\sigma_{22}^Z}{\sigma_{11}^Z\sigma_{22}^Z\sigma_{33}^Z} \\ 0 & \frac{1}{\sigma_{22}^Z} & -\frac{\sigma_{23}^Z}{\sigma_{22}^Z\sigma_{33}^Z} \\ 0 & 0 & \frac{1}{\sigma_{33}^Z} \end{bmatrix} \begin{bmatrix} 0 & -\lambda & -\lambda \\ 0 & \lambda & \lambda \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} \sigma_{11}^Z & \sigma_{12}^Z & \sigma_{13}^Z \\ 0 & \sigma_{22}^Z & \sigma_{23}^Z \\ 0 & 0 & \sigma_{33}^Z \end{bmatrix} \\ &= \begin{bmatrix} 0 & -\lambda\Big(\frac{\sigma_{12}^Z+\sigma_{22}^Z}{\sigma_{11}^Z}\Big) & -\lambda\Big(\frac{\sigma_{22}^Z(\sigma_{12}^Z+\sigma_{13}^Z+\sigma_{23}^Z+\sigma_{33}^Z)}{\sigma_{33}^Z\sigma_{22}^Z}\Big) \\ 0 & \lambda & \lambda\frac{\sigma_{33}^Z}{\sigma_{22}^Z} \\ 0 & 0 & \lambda \end{bmatrix}, \end{aligned} \tag{108}$$

which translates to implementing two supplementary parameter restrictions on the upper triangular mean-reversion matrix $\boldsymbol{K}_{\boldsymbol{Y}}^{\mathbb{Q}}$, i.e.

$$\kappa_{11}^{Y,\mathbb{Q}} = 0, \quad \kappa_{22}^{Y,\mathbb{Q}} = \kappa_{33}^{Y,\mathbb{Q}}. \tag{109}$$

The final restriction emerges from the short rate process as we lastly impose $\rho_0^Z = \rho_0^Y = 0$. Hence, the loading term $\boldsymbol{\rho_1^Y}$ remains unrestricted.

C Kalman Filter

C.1 State Filtering

We restate the linear Gaussian state space model via the formulation of Durbin & Koopman (2012) as

$$\boldsymbol{y_t} = \boldsymbol{d_t} + \boldsymbol{Z_t}\boldsymbol{\alpha_t} + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H_t}), \tag{110}$$

$$\boldsymbol{\alpha_{t+1}} = \boldsymbol{c_t} + \boldsymbol{T_t}\boldsymbol{\alpha_t} + \boldsymbol{\eta_t}, \qquad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q_t}), \qquad t = 1, \dots, T, \tag{111}$$

where $\boldsymbol{y_t}$ $(N \times 1)$ refers to the observed data at time t, $\boldsymbol{\alpha_t}$ $(J \times 1)$ coincides with the state vector at time t and $\Theta = \{\boldsymbol{d_t}$ $(N \times 1)$, $\boldsymbol{Z_t}$ $(N \times J)$, $\boldsymbol{H_t}$ $(N \times N)$, $\boldsymbol{c_t}$ $(J \times 1)$, $\boldsymbol{T_t}$ $(J \times J)$, $\boldsymbol{Q_t}$ $(J \times J)\}_{t=1}^T$ corresponds model's set of parameters respectively. Following Durbin & Koopman (2012) we refer to Equations (110) and (111) as the measurement equation and the state equation respectively.

The Kalman Filter (KF) extracts the signal of the states from the noise and comprises two intermediary steps, which we refer to as the update- and prediction step. Following Durbin & Koopman (2012) we define $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}} = \mathbb{E}[\boldsymbol{\alpha_t}|\mathcal{I}_t]$ as the best guess on the latent states at time t conditional on filtration $\mathcal{I}_t$ with uncertainties $\boldsymbol{P_{t|t}} = \text{Var}[\boldsymbol{\alpha_t}|\mathcal{I}_t]$. With these best guesses and uncertainties we conduct a prediction step of the KF to obtain an educated estimate of the

future states as

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}} = \boldsymbol{c_t} + \boldsymbol{T_t}\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}, \tag{112}$$

$$\boldsymbol{P}_{\boldsymbol{t+1|t}} = \boldsymbol{T_t}\boldsymbol{P}_{\boldsymbol{t|t}}\boldsymbol{T}_{\boldsymbol{t}}' + \boldsymbol{Q_t}, \tag{113}$$

where $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}} = \mathbb{E}[\boldsymbol{\alpha_{t+1}}|\mathcal{I}_t]$ and $\boldsymbol{P}_{\boldsymbol{t+1|t}} = \text{Var}[\boldsymbol{\alpha_{t+1}}|\mathcal{I}_t]$. As time passes by, we adjust our predictions in the update step via inference of the data. The update step exploits the joint normal distribution between the states and the data conditional on filtration $\mathcal{I}_t$, i.e.

$$\begin{pmatrix} \boldsymbol{y_{t+1}} \\ \boldsymbol{\alpha_{t+1}} \end{pmatrix} \Bigg| \boldsymbol{\Xi}, \mathcal{I}_t \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{Z_t}\hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}} \\ \hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}} \end{pmatrix}, \begin{bmatrix} \boldsymbol{Z_t}\boldsymbol{P}_{\boldsymbol{t+1|t}}\boldsymbol{Z}_{\boldsymbol{t}}' + \boldsymbol{H_t} & \boldsymbol{Z_t}\boldsymbol{P}_{\boldsymbol{t+1|t}} \\ \boldsymbol{P}_{\boldsymbol{t+1|t}}\boldsymbol{Z}_{\boldsymbol{t}}' & \boldsymbol{P}_{\boldsymbol{t+1|t}} \end{bmatrix} \right), \tag{114}$$

where we obtain the updates via the conditional normal lemma

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t+1}} = \hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}} + \boldsymbol{P}_{\boldsymbol{t+1|t}}\boldsymbol{Z}_{\boldsymbol{t}}'\boldsymbol{V}_{\boldsymbol{t}}^{-1}\boldsymbol{\vartheta_{t+1}}, \tag{115}$$

$$\boldsymbol{P}_{\boldsymbol{t+1|t+1}} = \boldsymbol{P}_{\boldsymbol{t+1|t}} - \boldsymbol{P}_{\boldsymbol{t+1|t}}\boldsymbol{Z}_{\boldsymbol{t}}'\boldsymbol{V}_{\boldsymbol{t}}^{-1}\boldsymbol{Z_t}\boldsymbol{P}_{\boldsymbol{t+1|t}}, \tag{116}$$

with $\boldsymbol{\vartheta_{t+1}} = \boldsymbol{y_{t+1}} - \boldsymbol{d_t} - \boldsymbol{Z_t}\hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}}$ and $\boldsymbol{V_t} = \boldsymbol{Z_t}\boldsymbol{P}_{\boldsymbol{t+1|t}}\boldsymbol{Z}_{\boldsymbol{t}}' + \boldsymbol{H_t}$. For a formal proof of the theorem we resort to Durbin & Koopman (2012). To initialize the filter we assume that $\boldsymbol{\alpha_1} \sim \mathcal{N}(\boldsymbol{a_1}, \boldsymbol{A_1})$ where both $\boldsymbol{a_1}$ and $\boldsymbol{A_1}$ are set by the researcher. If we suppress the time-dependence on the parameters in Equations (110) and (111), setting $\boldsymbol{a_1} = (\boldsymbol{I_p} - \boldsymbol{F})^{-1}\boldsymbol{c}$ and $\boldsymbol{A_1} = \text{vec}^{-1}\Big[\Big(\boldsymbol{I_{p^2}} - (\boldsymbol{F} \otimes \boldsymbol{F})\Big)^{-1}\text{vec}(\boldsymbol{Q})\Big]$ matches the unconditional moments (e.g. we refer to Hamilton et al. (1993) for a formal proof). The operation $\text{vec}(\cdot)$ coincides with vertically stacking a matrix' columns, whereas $\text{vec}^{-1}(\cdot)$ denotes its inverse operation.

C.2 State Smoothing

Bayesian inference requires the states to be conditioned, or equivalently smoothed, on all available information. Thus, we wish to derive the distribution of $\boldsymbol{\alpha_t}$ conditional on filtration $\mathcal{I}_T = \{\boldsymbol{y_t}\}_{t=1}^T$. Following Durbin & Koopman (2012) we refer to its conditional moments as $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|T}} = \mathbb{E}[\boldsymbol{\alpha_t}|\mathcal{I}_T]$ and $\boldsymbol{P}_{\boldsymbol{t|T}} = \text{Var}[\boldsymbol{P_t}|\mathcal{I}_T]$ $\forall t$. The joint density of $\boldsymbol{\alpha_t}$ and $\boldsymbol{\alpha_{t+1}}$ conditional on $\mathcal{I}_t$ yields backward recursions to obtain the smoothed states via

$$\begin{pmatrix} \boldsymbol{\alpha_t} \\ \boldsymbol{\alpha_{t+1}} \end{pmatrix} \Bigg| \boldsymbol{\Xi}, \mathcal{I}_T \sim \mathcal{N}\left[\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}} \\ \hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}} \end{pmatrix}, \begin{bmatrix} \boldsymbol{P}_{\boldsymbol{t|t}} & \boldsymbol{P}_{\boldsymbol{t+1|t}}\boldsymbol{T}_{\boldsymbol{t}}' \\ \boldsymbol{T_t}\boldsymbol{P}_{\boldsymbol{t|t}} & \boldsymbol{P}_{\boldsymbol{t+1|t}} \end{bmatrix} \right] \tag{117}$$

Following Durbin & Koopman (2012) we use the law of iterated probabilities to first condition on $\boldsymbol{\alpha_{t+1}}$ after which we average out all possible values of $\boldsymbol{\alpha_{t+1}}$. Hence, we depart from Equation (117) to derive

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|T}} = \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}} + \boldsymbol{P}_{\boldsymbol{t|t}}\boldsymbol{Z}_{\boldsymbol{t}}'\boldsymbol{P}_{\boldsymbol{t+1|t}}^{-1}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|T}} - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t+1|t}}), \tag{118}$$

where we construct the smoothed uncertainties via

$$\boldsymbol{P_{t|T}} = \boldsymbol{P_{t|t}} + \boldsymbol{P_{t|t}} \boldsymbol{Z_t'} \boldsymbol{P_{t+1|T}^{-1}} [\boldsymbol{P_{t+1|T}} - \boldsymbol{P_{t+1|t}}] \boldsymbol{P_{t+1|t}^{-1}} \boldsymbol{P_{t|t}}. \tag{119}$$

Following Durbin & Koopman (2012) we initialize the Kalman Smoother with the filtered states $\boldsymbol{\hat{\alpha}_{T|T}}$ and uncertainties $\boldsymbol{P_{T|T}}$ from Section C.1.

C.3 Prediction Error Decomposition

Given the filtered states from Section C.1, Christensen et al. (2011) and Ullah (2016) suggest to evaluate the log-likelihood function via the prediction error decomposition, i.e.

$$l(\boldsymbol{y_1}, \dots, \boldsymbol{y_t}|\Theta) = -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log(|\boldsymbol{V_t}|) + \boldsymbol{\vartheta_t'} \boldsymbol{V_t^{-1}} \boldsymbol{\vartheta_t}\right), \tag{120}$$

where $\boldsymbol{\vartheta_{t+1}} = \boldsymbol{y_{t+1}} - \boldsymbol{d_t} - \boldsymbol{Z_t} \boldsymbol{\hat{\alpha}_{t+1|t}}$ and $\boldsymbol{V_t} = \boldsymbol{Z_t} \boldsymbol{P_{t+1|t}} \boldsymbol{Z_t'} + \boldsymbol{H_t}$ (e.g. see Section C.1). Both Christensen et al. (2011) and Ullah (2016) employ the derivative-free Nelder-Mead method to maximize the likelihood with respect to Θ. Additionally, identical restrictions as to our Bayesian sampling routines apply here to ensure covariance stationarity (see Section 3.3 for an explanation in greater detail).

D Posterior Sampling Scheme for RNS

The state space representation of the RNS model from Section 2.2 comprises

$$\begin{aligned} \boldsymbol{y_t} &= \boldsymbol{B_\lambda^*} \boldsymbol{\beta}_t + \boldsymbol{\varepsilon_t}, & \boldsymbol{\varepsilon_t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \\ \boldsymbol{\beta_t} &= (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta_t}, & \boldsymbol{\eta_t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, \dots, T, \end{aligned}$$

where $\boldsymbol{B_\lambda^*}$ $(N \times 3)$ refers to the rotated factor loadings, $\boldsymbol{H}$ $(N \times N)$ corresponds to a diagonal covariance matrix, $\boldsymbol{\mu}$ (3×1) coincides with the unconditional means of the state equation, $\boldsymbol{F}$ (3×3) corresponds to the state dynamics matrix and $\boldsymbol{Q}$ (3×3) refers to the covariance matrix of the state equation. We exclude $\boldsymbol{B_\lambda^*}$ from the sampling scheme as we preliminary compute λ via the methods from Section 3 and hold it fixed throughout the entire sampling routine.

In this part of the appendix we construct the sampling schemes for two variants of the RNS model, i.e. the independent factor variant and the correlated factor variant, respectively. The former imposes an AR(1) process on the latent states and restricts both $\boldsymbol{F}$ and $\boldsymbol{Q}$ to be diagonal. Contrarily, the correlated factor variant fully parametrizes $\boldsymbol{F}$ and Cholesky decomposes $\boldsymbol{Q}$, i.e. $\boldsymbol{Q} = \boldsymbol{Q^{\frac{1}{2}}} \boldsymbol{Q^{\frac{1}{2}'}}$ via lower-triangular matrix $\boldsymbol{Q^{\frac{1}{2}}}$

D.1 Measurement Equation Sampling

The observation equation of the RNS model resolves to

$$\boldsymbol{y_t} = \boldsymbol{B_\lambda^*} \boldsymbol{\beta}_t + \boldsymbol{\varepsilon_t}, \quad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}),$$

which is equivalent for both variants of the RNS model. We assume the disturbances to be white noise processes, such that $\boldsymbol{H} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$. Hence, we sample the i-th diagonal component of $\boldsymbol{H}$ via univariate formulation of the measurement equation as

$$\begin{aligned}
&\Leftrightarrow y_t(\tau_i) = \beta_{1,t} + \beta_{2,t}\left[1 - \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}\right] + \beta_{3,t}\left[\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right] + \varepsilon_{i,t}, \\
&\Leftrightarrow y_{i,t} \quad = \boldsymbol{b_i'\beta_t} + \varepsilon_{i,t}
\end{aligned}$$

where $\boldsymbol{\beta_t} = [\beta_{1,t}, \beta_{2,t}, \beta_{3,t}]'$ and $\boldsymbol{b_i} = [1, 1 - \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}, \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}]'$ for $i = 1, 2, \ldots, N$. In case we implement an independent $IG2(\underline{\sigma}_i^2, \underline{\nu}_{\sigma_i})$ prior, the posterior sampling distribution of σ_i^2 resolves to an inverted gamma-2 distribution with shape parameter $\underline{\sigma}_i^2 + \sum_{t=1}^{T}\left(y_t(\tau_i) - \boldsymbol{b_i'\beta_t}\right)^2$ and $T + \underline{\nu}_{\sigma_i}$ degrees of freedom.

D.2 State Equation Sampling

The difference in sampling schemes for both RNS models lies in the state equations. We elaborate on the univariate methods used to sample the state equation of the independent factor RNS model in Appendix D.2.1, whereas we provide the multivariate full conditional marginal posteriors for the correlated factor RNS model in Appendix D.2.2.

D.2.1 Independent Factor RNS

For the independent factor variant of the RNS model we subdivide the state equation in its univariate components, i.e.

$$\beta_{j,t} = (1 - f_{jj})\mu_j + f_{jj}\beta_{j,t-1} + \eta_{j,t}, \quad j = 1, 2, 3, \quad t = 2, \ldots T,$$

where $\beta_{j,t}$ refers to the j-th factor at time t, μ_j coincides with the j-th element of $\boldsymbol{\mu}$, f_{jj} and q_{jj}^2 correspond to the j-th diagonal elements of $\boldsymbol{F}$ and $\boldsymbol{Q}$ respectively and $\eta_{j,t}$ denotes the j-th element of the state disturbances with $\eta_{j,t} \sim \mathcal{N}(0, q_{jj}^2)$.

Following the order from Section 3, we first derive the posterior sampling distributions for the covariances on the state equation. For an inverse gamma-2 prior with shape parameter $\underline{q}_{jj}^2$ and degrees of freedom $\underline{\nu}_{q_{jj}}$, we draw q_{jj}^2 from an $IG2\left(\bar{q}_{jj}^2, \bar{\nu}_{q_{jj}}\right)$ with shape parameter

$$\bar{q}_{jj}^2 = \underline{q}_{jj}^2 + \sum_{t=2}^{T}[(\beta_{j,t} - \mu_j) - f_{jj}(\beta_{j,t-1} - \mu_j)]^2$$

and $\bar{\nu}_{q_{jj}} = (T-1) + \underline{\nu}_{q_{jj}}$ degrees of freedom.

In order to derive the posterior sampling distributions for the state dynamics, we reformulate

the univariate representation of the state equation as

$$\begin{aligned}
&\Leftrightarrow \beta_{j,t} &&= \mu_j(1 - f_{jj}) + f_{jj}\beta_{j,t-1} + \eta_{j,t},\\
&\Leftrightarrow \beta_{j,t} - f_{jj}\beta_{j,t-1} &&= \mu_j(1 - f_{jj}) + \eta_{j,t},\\
&\Leftrightarrow \beta_{j,t} - \mu_j &&= f_{jj}(\beta_{j,t-1} - \mu_j) + \eta_{j,t},
\end{aligned}$$

where all three expressions are equivalent representations of the same regression. We consider independent normal priors for both μ_j and f_{jj}, i.e. $\mu_j \sim \mathcal{N}(\underline{\mu}_j, \underline{V}_{\mu_j})$ and $f_{jj} \sim \mathcal{N}(\underline{f}_{jj}, \underline{V}_{f_{jj}})$, respectively. Then, we construct the posterior distribution of $\mu_j \sim \mathcal{N}(\bar{\mu}_j, \bar{V}_{\mu_j})$ via synthetically created regressors $x_{j,t} = 1 - f_{jj}$ and $y_{j,t} = \beta_{j,t} - f_{jj}\beta_{j,t-1}$ and independent conjugate $\mathcal{N}(\underline{\mu}_j, \underline{V}_{\mu_j})$, where

$$\bar{\mu}_j = \left[\sum_{t=2}^{T}\left(\frac{x_{j,t}}{q_{jj}}\right)^2 + \underline{V}_{\mu_j}^{-1}\right]^{-1}\left[\sum_{t=2}^{T}\left(\frac{x_{j,t}y_{j,t}}{q_{jj}^2}\right) + \frac{\underline{\mu}_j}{\underline{V}_{\mu_j}}\right]$$

and

$$\bar{V}_{\mu_j} = \left[\sum_{t=2}^{T}\left(\frac{x_{j,t}}{q_{jj}}\right)^2 + \underline{V}_{\mu_j}^{-1}\right]^{-1},$$

for $j = 1, 2, 3$. Similarly, we reformulate the regressors as $x_{j,t} = \beta_{j,t-1} - \mu_j$ and $y_{j,t} = \beta_{j,t} - \mu_j$ to generate the posterior sampling distribution for the j-th autoregressive coefficient of the state equation. Hence, we write that $f_{jj} \sim \mathcal{N}(\bar{f}_{jj}, \bar{V}_{f_{jj}})$ with

$$\bar{f}_{jj} = \left[\sum_{t=2}^{T}\left(\frac{x_{j,t}}{q_{jj}}\right)^2 + \underline{V}_{f_{jj}}^{-1}\right]^{-1}\left[\sum_{t=2}^{T}\left(\frac{x_{j,t}y_{j,t}}{q_{jj}^2}\right) + \frac{\underline{f}_{jj}}{\underline{V}_{f_{jj}}}\right]$$

and

$$\bar{V}_{f_{jj}} = \left[\sum_{t=2}^{T}\left(\frac{x_{j,t}}{q_{jj}}\right)^2 + \underline{V}_{f_{jj}}^{-1}\right]^{-1},$$

in case we implement an independent conjugate $\mathcal{N}(\underline{f}_{jj}, \underline{V}_{f_{jj}})$ prior. Moreover, we assure the state equation of the independent factor RNS model to be stationary via truncated normal sampling, i.e. we follow Hautsch & Yang (2012) and restrict the acquired samples via $0 < |f_{jj}| < 1$ for $j = 1, 2, 3$.

D.2.2 Correlated Factor RNS

The sampling routine for the correlated factor variant of the RNS model encompasses posterior inference via multivariate distributions, although the pattern shows great similarity to the independent factor RNS model. We recall the state equation of the correlated factor RNS model as

$$\boldsymbol{\beta_t} = (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta_{t-1}} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}),$$

where we first focus on the full conditional marginal posterior density of covariance matrix $\boldsymbol{Q}$. For an independent inverted Wishart prior with shape parameter $\underline{\boldsymbol{Q}}$ and $\underline{\nu}_Q$ degrees of freedom, we obtain an inverted Wishart posterior density with shape parameter $\bar{\boldsymbol{Q}} = \underline{\boldsymbol{Q}} + \sum_{t=2}^{T}(\boldsymbol{\beta_t} - \boldsymbol{\mu} -$

$\boldsymbol{F}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu}))(\boldsymbol{\beta_t} - \boldsymbol{\mu} - \boldsymbol{F}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu}))'$ and $\bar{\nu}_Q = (T-1) + \underline{\nu}_Q$ degrees of freedom.

In line with the derivation of the marginal posteriors for the independent factor RNS model, reformulation of the state equation yields

$$\begin{aligned}
\Leftrightarrow \quad & \boldsymbol{\beta_t} && = (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta_{t-1}} + \boldsymbol{\eta_t} \\
\Leftrightarrow \quad & \boldsymbol{\beta_t} - \boldsymbol{F}\boldsymbol{\beta_{t-1}} && = (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \\
\Leftrightarrow \quad & \boldsymbol{\beta_t} - \boldsymbol{\mu} && = \boldsymbol{F}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu}) + \boldsymbol{\eta_t},
\end{aligned}$$

where all three expressions are equivalent representation of the same regression, such that we extend the results from Appendix D.2.1 to the correlated factor variant of the RNS model.

In order to derive the multivariate sampling distribution for unconditional mean vector $\boldsymbol{\mu}$, we first ease notation via $\boldsymbol{X} = \boldsymbol{I_3} - \boldsymbol{F}$ and $\boldsymbol{z_t} = \boldsymbol{\beta_t} - \boldsymbol{F}\boldsymbol{\beta_{t-1}}$. Then, we standardize the state equation to obtain a multivariate normal posterior for $\boldsymbol{\mu}$ with mean $\bar{\boldsymbol{\mu}}$ and variance $\bar{\boldsymbol{V}}_{\boldsymbol{\mu}}$, where

$$\bar{\boldsymbol{\mu}} = \left[(T-1)\boldsymbol{X}'\boldsymbol{Q}^{-1}\boldsymbol{X} + \underline{\boldsymbol{V}}_{\boldsymbol{\mu}}^{-1}\right]^{-1}\left[\left(\sum_{t=2}^{T}\boldsymbol{X}'\boldsymbol{Q}^{-1}\boldsymbol{z_t}\right) + \underline{\boldsymbol{V}}_{\boldsymbol{\mu}}^{-1}\underline{\boldsymbol{\mu}}\right]$$

and

$$\bar{\boldsymbol{V}}_{\boldsymbol{\mu}} = \left[(T-1)\boldsymbol{X}'\boldsymbol{Q}^{-1}\boldsymbol{X} + \underline{\boldsymbol{V}}_{\boldsymbol{\mu}}^{-1}\right]^{-1},$$

in case we implement an independent $\mathcal{N}(\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{V}}_{\boldsymbol{\mu}})$ prior. Next, we follow Çakmaklı (2020) to vectorize $\boldsymbol{F}$ for more convenient prior intervention. We synthetically compose regressors $\boldsymbol{z_t} = \boldsymbol{\beta_t} - \boldsymbol{\mu}$ and $\boldsymbol{x_t} = \boldsymbol{\beta_{t-1}} - \boldsymbol{\mu}$, such that reformulation of the regression in compact form yields

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{X} + \boldsymbol{U},$$

where $\boldsymbol{Z} = [\boldsymbol{z_1}, \boldsymbol{z_2}, \dots \boldsymbol{z_T}]$, $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_T}]$ and $\boldsymbol{U} = [\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \dots \boldsymbol{\eta_T}]$ with $\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{U} \in \mathbb{R}^{J \times T}$. It follows that $\text{vec}(\boldsymbol{Z} - \boldsymbol{F}\boldsymbol{X}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = \boldsymbol{Q} \otimes \boldsymbol{I_{T-1}}$. Hence, we obtain a $\mathcal{N}(\bar{\boldsymbol{\mu}}_{\text{vec}(\boldsymbol{F})}, \bar{\boldsymbol{V}}_{\text{vec}(\boldsymbol{F})})$ posterior with

$$\begin{aligned}
\bar{\boldsymbol{V}}_{\text{vec}(\boldsymbol{F})} &= \left(\underline{\boldsymbol{V}}_{\text{vec}(\boldsymbol{F})} + (\boldsymbol{X}' \otimes \boldsymbol{I_3})'\boldsymbol{\Omega}^{-1}(\boldsymbol{X}' \otimes \boldsymbol{I_3})\right)^{-1}, \\
\bar{\boldsymbol{\mu}}_{\text{vec}(\boldsymbol{F})} &= \bar{\boldsymbol{V}}_{\text{vec}(\boldsymbol{F})}\left(\underline{\boldsymbol{V}}_{\text{vec}(\boldsymbol{F})}\underline{\boldsymbol{\mu}}_{\text{vec}(\boldsymbol{F})} + (\boldsymbol{X}' \otimes \boldsymbol{I_3})'\boldsymbol{\Omega}^{-1}(\boldsymbol{X}' \otimes \boldsymbol{I_3})\text{vec}(\boldsymbol{F})_{GLS},\right.
\end{aligned}$$

where

$$\text{vec}(\boldsymbol{F})_{GLS} = \left((\boldsymbol{X}' \otimes \boldsymbol{I_3})'\boldsymbol{\Omega}^{-1}(\boldsymbol{X}' \otimes \boldsymbol{I_3})\right)^{-1}\left(\boldsymbol{X}' \otimes \boldsymbol{I_3})'\boldsymbol{\Omega}^{-1}\text{vec}(\boldsymbol{Z})\right),$$

if we opt for a multivariate $\mathcal{N}(\underline{\boldsymbol{\mu}}_{\text{vec}(\boldsymbol{F})}, \underline{\boldsymbol{V}}_{\text{vec}(\boldsymbol{F})})$. Following Christensen et al. (2011), we ensure the state equation to be covariance stationary via rejection sampling on $\boldsymbol{F}$. As such, we reject samples of $\boldsymbol{F}$ with eigenvalues larger than one.

D.3 State Sampling

We account for uncertainty in the smoothed states via the second algorithm of Durbin & Koopman (2002). They provide an efficient simulation smoother that only requires to run both the Kalman Filter and Kalman Smoother once per iteration. Following our state space representation of the RNS model as

$$\begin{aligned} \boldsymbol{y_t} &= \boldsymbol{B^*_\lambda}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon_t}, & \boldsymbol{\varepsilon_t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \\ \boldsymbol{\beta_t} &= (\boldsymbol{I_3} - \boldsymbol{F})\boldsymbol{\mu} + \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta_t}, & \boldsymbol{\eta_t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, \dots, T, \end{aligned}$$

we follow the notation from Durbin & Koopman (2002) and write $\boldsymbol{\beta} = (\boldsymbol{\beta'_1}, \boldsymbol{\beta'_2}, \dots, \boldsymbol{\beta'_T})$, $\boldsymbol{w} = (\boldsymbol{\varepsilon'_1}, \boldsymbol{\varepsilon'_2}, \dots \boldsymbol{\varepsilon'_T}, \boldsymbol{\eta'_1}, \boldsymbol{\eta'_2}, \dots, \boldsymbol{\eta'_T})'$ and $\boldsymbol{y} = (\boldsymbol{y'_1}, \boldsymbol{y'_2}, \dots \boldsymbol{y'_T})'$. The goal of the smoother is to obtain smoothed states $\boldsymbol{\beta}$ from $p(\boldsymbol{\beta}|\boldsymbol{y})$. To acquire the desired posterior results, we define $\boldsymbol{w}^+$ as a random draw from $p(\boldsymbol{w})$, $\boldsymbol{\beta}^+$ as a random draw from $p(\boldsymbol{\beta})$ and $\tilde{\boldsymbol{\beta}}$ as the desired draw from $p(\boldsymbol{\beta}|\boldsymbol{y})$. Then, the simulation smoother corresponds to the following procedure:

1. Draw $\boldsymbol{w}^+$ from $p(\boldsymbol{w}) = \mathcal{N}\Big(\boldsymbol{0}, \text{diag}[\boldsymbol{H} \otimes \boldsymbol{I_{T-1}}, \boldsymbol{Q} \otimes \boldsymbol{I_{T-1}}]\Big)$.
2. Draw $\boldsymbol{\beta_1} \sim \mathcal{N}(\boldsymbol{b_1}, \boldsymbol{B_1})$ and generate $\boldsymbol{y}^+$ and $\boldsymbol{\beta}^+$ by jointly iterating the state equation and the measurement equation forward while we account for stochastic uncertainty via the sampling of the respective disturbances. We diffusely initialize the procedure and set $\boldsymbol{b_1} = \boldsymbol{\mu}$ and $\boldsymbol{B_1} = 10^6 \times \boldsymbol{I_3}$
3. Create artificial observations $\boldsymbol{y^*} = \boldsymbol{y} - \boldsymbol{y}^+$ and use these to run both the Kalman Filter and Kalman Smoother to obtain $\hat{\boldsymbol{\beta}}^* = \mathbb{E}[\boldsymbol{\beta}|\boldsymbol{y^*}]$ (e.g. see Appendix C).
4. Compute the posterior draw of the states via $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^* + \boldsymbol{\beta}^+$.

E Posterior Sampling Scheme AFRNS

The state space composition of the AFRNS model from Section 2.3 corresponds to

$$\begin{aligned} \boldsymbol{y_t} &= \boldsymbol{a} + \boldsymbol{B^*_\lambda}\boldsymbol{\beta_t} + \boldsymbol{\varepsilon_t}, & \boldsymbol{\varepsilon_t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \\ \boldsymbol{\beta_t} &= \Big(\boldsymbol{I} - \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}\Big)\boldsymbol{\theta}^{\mathbb{P}} + \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}\Delta t}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta_t}, & \boldsymbol{\eta_t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, \dots, T, \end{aligned}$$

where $\boldsymbol{a}$ $(N \times 1)$ denotes the yield-adjustment, $\boldsymbol{B^*_\lambda}$ $(N \times 3)$ refers to the rotated factor loadings, $\boldsymbol{H}$ $(N \times N)$ corresponds to a diagonal covariance matrix, $\boldsymbol{\theta}^{\mathbb{P}}$ (3×1) coincides with the process' drift under the $\mathbb{P}$-measure, $\boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}}}$ (3×3) corresponds to the discretized state dynamics and $\boldsymbol{Q}$ (3×3) refers to the covariance matrix on the state equation. Our Bayesian approach, which inversely ties the RNS model to the framework from Duffie & Kan (1996), is limited to an independent factor variant of the AFRNS model (e.g. see Section 3 for an explanation in greater detail). Hence, there is no sampling scheme for the correlated factor variant of the AFRNS model presented in this part of the appendix.

E.1 Measurement Equation Sampling

We sample σ_i^2 from an inverted gamma-2 density with shape parameter $\underline{\varrho}_i^2 + \sum_{t=1}^{T} (y_{i,t} - a_i - \boldsymbol{b}_{i,t}'\boldsymbol{\beta}_t)^2$ and $T + \bar{\nu}_\sigma$ degrees of freedom if we use an independent $IG2(\underline{\varrho}^2, \underline{\nu}_\sigma)$ prior, for $i = 1, 2, \ldots, N$ (e.g. see Appendix D).

We cannot randomly draw the yield-adjustments a_i, $i = 1, 2, ..., N$, as they resolve to deterministic functions conditional on $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ and τ_i. Hence, we filter the risk-neutral volatilities via the methods in Appendix E.2 after which we compute the yield-adjustment via Appendix A.3.

E.2 State Equation Sampling

The sampling routine of the independent factor AFRNS model yields great similarity to Appendix D.2.1, although the extraction of the continuous-time parameters from Equation (20) requires some extra steps.

Following Appendix D.2.1, we subdivide the state equation of the independent factor AFRNS model in its univariate components, i.e.

$$\beta_{j,t} = \left(1 - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\right)\theta_j^{\mathbb{P}} + e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\beta_{j,t-1} + \eta_{j,t}, \quad j = 1, 2, 3, \quad t = 2, \ldots T,$$

where $\beta_{j,t}$ refers to the j-th factor at time t, $\theta_j^{\mathbb{P}}$ coincides with the j-th element of $\boldsymbol{\theta}^{\mathbb{P}}$, $\kappa_{jj}^{\mathbb{P}}$ denotes the j-th diagonal element of continuous-time state dynamics matrix $\boldsymbol{K}^{\mathbb{P}}$ and $\eta_{j,t}$ corresponds to the j-th element of the state disturbances with $\eta_{j,t} \sim \mathcal{N}(0, q_{jj}^2)$. The posterior sampling distribution of q_{jj}^2 resolves to an inverse gamma-2 distribution with $\bar{\nu}_{q_{jj}} = (T-1) + \underline{\nu}_{q_{jj}}$ degrees of freedom and shape parameter $\bar{q}_j^2 = \underline{q}_j^2 + \sum_{t=1}^{T} \left(\beta_{j,t} - \left(1 - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\right)\theta_j^{\mathbb{P}} - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\beta_{j,t-1}\right)^2$.

Next, we derive the posterior distributions for the parameters on the j-th state equation via three equivalent representations of the same regression as

$$\begin{aligned}
\Leftrightarrow \quad & \beta_{j,t} && = \left(1 - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\right)\theta_j^{\mathbb{P}} + e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\beta_{j,t-1} + \eta_{j,t}, \\
\Leftrightarrow \quad & \beta_{j,t} - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\beta_{j,t-1} && = \left(1 - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\right)\theta_j^{\mathbb{P}} + \eta_{j,t}, \\
\Leftrightarrow \quad & \beta_{j,t} - \theta_j^{\mathbb{P}} && = (\beta_{j,t-1} - \theta_j^{\mathbb{P}})f_{jj} + \eta_{j,t}.
\end{aligned}$$

Firstly, we focus on the posterior sampling distribution of the j-th drift component $\theta_j^{\mathbb{P}}$. In line with Appendix D.2.1, we synthetically construct regressors $y_{j,t} = \beta_{j,t} - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}\beta_{j,t-1}$ and $x_{j,t} = (1 - e^{-\kappa_{jj}^{\mathbb{P}}\Delta t})$ to further ease notation. For an independent $\mathcal{N}(\underline{\theta}_j^{\mathbb{P}}, \underline{V}_{\theta_j^{\mathbb{P}}})$ prior, the posterior distribution of $\theta_j^{\mathbb{P}}$ resolves to a $\mathcal{N}(\bar{\theta}_j^{\mathbb{P}}, \bar{V}_{\theta_j^{\mathbb{P}}})$ density where

$$\bar{\theta}_j^{\mathbb{P}} = \left[\sum_{t=2}^{T}\left(\frac{x_{j,t}}{q_{jj}}\right)^2 + \underline{V}_{\theta_j^{\mathbb{P}}}^{-1}\right]^{-1}\left[\sum_{t=2}^{T}\left(\frac{x_{j,t}y_{j,t}}{q_{jj}^2}\right) + \frac{\underline{\theta}_j^{\mathbb{P}}}{\underline{V}_{\theta_j^{\mathbb{P}}}}\right],$$

and

$$\bar{V}_{\theta_j^{\mathbb{P}}} = \left[\sum_{t=2}^{T} \left(\frac{x_{j,t}}{q_{jj}} \right)^2 + \underline{V}_{\theta_j^{\mathbb{P}}}^{-1} \right]^{-1}.$$

Difficulties arise when one seeks to derive the Gibbs sampler for $e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}$ as it does not fit a standard regression framework. However, we exploit the unrestricted $\mathbb{P}$-dynamics of the AFRNS model to treat $e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}$ as a regular parameter. Hence, we follow the procedures in Appendix D.2.1 and introduce $f_{jj} = e^{-\kappa_{jj}^{\mathbb{P}}\Delta t}$ as the discretized autoregressive parameter. Subsequently, we compute its continuous-time counterpart via $\kappa_{jj}^{\mathbb{P}} = -\log(f_{jj})/\Delta t$. Additionally, we follow Christensen et al. (2011) and ensure the state equation to be stationary via non-zero truncated normal sampling of the continuous-time state dynamics, i.e. $\kappa_{jj} > 0$ for $j = 1, 2, 3$.

As the limited structure of $\boldsymbol{Q}$ ties the continuous-time $\mathbb{P}$-dynamics to the risk-neutral volatilities, we derive an algebraic expression for the latter via inverse formulation of Equation (28), i.e.

$$\boldsymbol{Q} = \int_0^{\Delta t} \boldsymbol{e}^{-\boldsymbol{K}^{\mathbb{P}} s} \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \boldsymbol{e}^{-(\boldsymbol{K}^{\mathbb{P}})' s} \mathrm{d}s,$$

$$\begin{bmatrix} q_{11}^2 & 0 & 0 \\ 0 & q_{22}^2 & 0 \\ 0 & 0 & q_{33}^2 \end{bmatrix} = \int_0^{\Delta t} \begin{bmatrix} \sigma_{11}^2 e^{-\kappa_{11}^{\mathbb{P}} s} & 0 & 0 \\ 0 & \sigma_{22}^2 e^{-\kappa_{22}^{\mathbb{P}} s} & 0 \\ 0 & 0 & \sigma_{33}^2 e^{-\kappa_{33}^{\mathbb{P}} s} \end{bmatrix} \mathrm{d}s,$$

where the risk-neutral volatilities resolve to deterministic functions of the freely-sampled covariances q_{jj}^2 and regressors $\kappa_{jj}^{\mathbb{P}}$. Hence, we obtain for the j-th diagonal element of $\boldsymbol{\Sigma}$ that $\sigma_{jj} = \sqrt{\left(2q_{jj}^2 \kappa_{jj}^{\mathbb{P}}\right) / \left(1 - e^{-2\kappa_{jj}^{\mathbb{P}}\Delta t}\right)}$.

E.3 State Sampling

We sample the states of the AFRNS model with the procedure from Appendix D, only to find an additional constant in the measurement equation. Hence, we refer to Appendix D for an explanation in greater detail.

F Posterior Sampling Scheme MF-RNS

The Gaussian state space formulation of Macro-Finance Rotated Nelson-Siegel (MF-RNS) resolves to

$$\begin{pmatrix} \boldsymbol{y_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \begin{bmatrix} \boldsymbol{B_\lambda^*} & \boldsymbol{0_{N\times P}} \\ \boldsymbol{0_{P\times 3}} & \boldsymbol{I_P} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon_t} \\ \boldsymbol{0_P} \end{pmatrix}, \quad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}),$$

$$\begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \left(\boldsymbol{I_{3+P}} - \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \right) \begin{pmatrix} \boldsymbol{\mu_\beta} \\ \boldsymbol{\mu_\gamma} \end{pmatrix} + \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_{t-1}} \\ \boldsymbol{\gamma_{t-1}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix},$$

where $\boldsymbol{\gamma_t}$ $(P \times 1)$ encompasses the exogenous factors at time t, $\boldsymbol{B^*_\lambda}$ $(N \times 3)$ coincides with the rotated factor loadings, $\boldsymbol{H}$ $(N \times N)$ refers to a diagonal covariance matrix, matrices $\boldsymbol{K_1}$ (3×3), $\boldsymbol{K_2}$ $(3 \times P)$, $\boldsymbol{K_3}$ $(P \times 3)$ and $\boldsymbol{K_4}$ $(P \times P)$ refer to the merged state dynamics between the latent factors and observable series, $\boldsymbol{\mu_\beta}$ (3×1) coincides with the unconditional means of the latent states and $\boldsymbol{\mu_\gamma}$ $(P \times 1)$ refers to the unconditional means of the exogenous factors. The assumed error structure for the state equation of the MF-RNS model is

$$\begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \right),$$

where $\boldsymbol{Q}$ (3×3) corresponds to the covariance matrix of the latent factors and $\boldsymbol{R}$ $(P \times P)$ refers to the covariance matrix of the observable series. A similar composition of the disturbances on the state equation ensures the VAR(1) processes of the latent factors and the observable series to be orthogonal. An independent factor variant of the MF-RNS imposes a diagonal structure on both $\boldsymbol{K_1}$ and $\boldsymbol{Q}$, although it leaves $\boldsymbol{K_2}$, $\boldsymbol{K_3}$ and $\boldsymbol{K_4}$ untouched. In contrast, a correlated factor variant of the MF-RNS model allows for a fully-flexible matrix $\boldsymbol{K_1}$ and a Cholesky-decomposable matrix $\boldsymbol{Q}$. In line with Appendices D and E we exclude factor loading matrix $\boldsymbol{B^*_\lambda}$ from the posterior sampling scheme as its completion is deterministic conditional on λ (Diebold & Li (2006)).

F.1 Measurement Equation Sampling

The sampling of $\{\sigma_i^2\}_{i=1}^N$ yields identical procedures for the RNS model and the MF-RNS model, respectively. Hence, we refer to Appendix D for an explanation in greater detail.

F.2 State Equation Sampling

As the independent factor variant of the MF-RNS only imposes a different structure on the matrices $\boldsymbol{K_1}$ and $\boldsymbol{Q}$, we obtain some commonalities in the respective sampling schemes of the state equations for both variants of the MF-RNS model. Accordingly, we derive the posterior distribution of $\boldsymbol{\mu} = [\boldsymbol{\mu'_\beta}, \boldsymbol{\mu'_\gamma}]'$ via reformulation of the state equation as

$$\boldsymbol{\xi_t} = (\boldsymbol{I_{3+J}} - \boldsymbol{K})\boldsymbol{\mu} + \boldsymbol{K}\boldsymbol{\xi_{t-1}} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}),$$

where

$$\boldsymbol{\xi_t} = \begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_\beta} \\ \boldsymbol{\mu_\gamma} \end{pmatrix}, \quad \boldsymbol{\epsilon_t} = \begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix}, \quad \boldsymbol{K} = \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix}.$$

which yields similarities to the posterior sampling scheme of $\boldsymbol{\mu}$ from the correlated RNS model. Hence, we rephrase the state equation as

$$\Leftrightarrow \qquad \boldsymbol{\xi_t} = (\boldsymbol{I_{3+J}} - \boldsymbol{K})\boldsymbol{\mu} + \boldsymbol{K}\boldsymbol{z_{t-1}} + \boldsymbol{\epsilon_t}, \Leftrightarrow \qquad \boldsymbol{\xi_t} - \boldsymbol{K}\boldsymbol{\xi_{t-1}} = (\boldsymbol{I_{3+J}} - \boldsymbol{K})\boldsymbol{\mu} + \boldsymbol{\epsilon_t},$$

where we simplify notation via $\boldsymbol{X} = \boldsymbol{I_{3+J}} - \boldsymbol{K}$ and $\boldsymbol{z_t} = \boldsymbol{\xi_t} - \boldsymbol{K}\boldsymbol{\xi_{t-1}}$. Thus, we draw $\boldsymbol{\mu}$ from a multivariate normal distribution with mean

$$\bar{\boldsymbol{\mu}} = \left[(T-1)\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X} + \underline{\boldsymbol{V}}_{\boldsymbol{\mu}}^{-1}\right]^{-1}\left[\left(\sum_{t=1}^{T}\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{z_t}\right) + \underline{\boldsymbol{V}}_{\boldsymbol{\mu}}^{-1}\underline{\boldsymbol{\mu}}\right]$$

and variance

$$\bar{\boldsymbol{V}}_{\boldsymbol{\mu}} = \left[\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X} + \underline{\boldsymbol{V}}_{\boldsymbol{\mu}}^{-1}\right]^{-1},$$

in case we use an independent multivariate normal prior for $\boldsymbol{\mu}$ with mean $\underline{\boldsymbol{\mu}}$ and variance $\underline{\boldsymbol{V}}_{\boldsymbol{\mu}}$.

Both variants of the MF-RNS model offer greater flexibility on covariance matrix $\boldsymbol{R}$, i.e. both do not assume a diagonal structure on $\boldsymbol{R}$. Hence, we derive its posterior distribution via reformulation of the state equation's lower partition as

$$\boldsymbol{\gamma_t} = \boldsymbol{\mu_\gamma} + \boldsymbol{K_3}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta}) + \boldsymbol{K_4}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) + \boldsymbol{\omega_t}, \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}),$$

where the posterior of $\boldsymbol{R}$ comprises an inverted Wishart distribution with shape parameter

$$\bar{\boldsymbol{R}} = \sum_{t=2}^{T}\left((\boldsymbol{\gamma_t} - \boldsymbol{\mu_\gamma}) - \begin{bmatrix}\boldsymbol{K_3} & \boldsymbol{K_4}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta} \\ \boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}\end{bmatrix}\right)\left((\boldsymbol{\gamma_t} - \boldsymbol{\mu_\gamma}) - \begin{bmatrix}\boldsymbol{K_3} & \boldsymbol{K_4}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta} \\ \boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}\end{bmatrix}\right)' + \underline{\boldsymbol{R}}$$

and $(T-1) + \nu_R$ degrees of freedom in case we use an independent $IW(\underline{\boldsymbol{R}}, \nu_R)$ prior.

F.2.1 Independent Factor MF-RNS

The independent factor MF-RNS model allows the latent factors to readily interfere with the set of exogenous factors. This results in a specific structure of the state dynamics matrices, i.e. we obtain a diagonal matrix $\boldsymbol{K_1}$ and a fully-flexible structure for $\boldsymbol{K_2}$, $\boldsymbol{K_3}$ and $\boldsymbol{K_4}$. In order to derive the posterior results for the state dynamics matrices, we distinguish four steps. First, write for the j-th latent factor that

$$\begin{aligned}
&\Leftrightarrow \quad \beta_{j,t} - \mu_{\beta,j} &&= \kappa_{jj}^1(\beta_{j,t-1} - \mu_{\beta,j}) + \boldsymbol{\kappa_{j\cdot}^2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) + \eta_{j,t} \\
&\Leftrightarrow \quad \beta_{j,t} - \mu_{\beta,j} - \boldsymbol{\kappa_{j\cdot}^2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) &&= \kappa_{jj}^1(\beta_{j,t-1} - \mu_{\beta,j}) + \eta_{j,t}, \quad \eta_{j,t} \sim \mathcal{N}(0, q_{jj}^2), \quad j \in \{1,2,3\},
\end{aligned}$$

where $\beta_{j,t}$ refers to the j-th factor at time t, $\mu_{\beta,j}$ coincides with the unconditional mean of the j-th latent factor, κ_{jj}^1 corresponds to the j-th diagonal autoregressive coefficient in $\boldsymbol{K_1}$, $\eta_{j,t}$ refers to the j-th disturbance on the state equation such that $\eta_{j,t} \sim \mathcal{N}(0, q_{jj}^2)$, q_{jj}^2 correspond the j-th diagonal covariance from matrix $\boldsymbol{Q}$ and $\boldsymbol{k_{j\cdot}^2}$ refers to the j-th row of $\boldsymbol{K_2}$. We sample κ_{jj}^1 via a standard regression result. For notational ease, we construct regressors $x_{j,t} = \beta_{j,t-1} - \mu_{\beta,j}$ and $y_{j,t} = \beta_{j,t} - \mu_{\beta,j} - \boldsymbol{k_{j\cdot}^2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma})$.

The posterior distribution of q_{jj}^2 corresponds to an inverse gamma-2 distribution with $(T -$

$1) + \nu_{q_{jj}}$ degrees of freedom and shape parameter

$$\bar{q}_j^2 = \sum_{t=2}^{T} \left((\beta_{j,t} - \mu_{\beta,j}) - k_{jj}^1 (\beta_{j,t-1} - \mu_{\beta,j}) - \boldsymbol{k_{j\cdot}^2} (\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) \right)^2 + \underline{q}_j^2,$$

in case we implement an independent $IG2(\underline{q}_j^2, \nu_{q_{jj}})$ prior. Secondly, we obtain a univariate normal posterior distribution for κ_{jj}^1 with mean

$$\bar{k}_{jj}^1 = \left[\sum_{t=2}^{T} \left(\frac{x_{j,t}}{q_{jj}} \right)^2 + \underline{V}_{k_{jj}^1}^{-1} \right]^{-1} \left[\sum_{t=2}^{T} \left(\frac{x_{j,t} y_{j,t}}{q_{jj}^2} \right) + \frac{\underline{k}_{jj}^1}{\underline{V}_{k_{jj}^1}} \right],$$

and variance

$$\bar{V}_{k_{jj}^1} = \left[\sum_{t=2}^{T} \left(\frac{x_{j,t}}{q_{jj}} \right)^2 + \underline{V}_{k_{jj}^1}^{-1} \right]^{-1},$$

if we opt for an $\mathcal{N}(\underline{\kappa}_{jj}^1, \underline{V}_{\kappa_{jj}^1})$ prior. Chronologically, we rephrase the state equation's upper partition as

$$\boldsymbol{\beta_t} - \boldsymbol{\mu_\beta} - \boldsymbol{K_1}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta}) = \boldsymbol{K_2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}),$$

which allows us to assess $\boldsymbol{K_2}$'s posterior distribution using standard multivariate regression results. Similarly, we reformulate the state equation's lower partition as

$$\begin{aligned} &\Leftrightarrow \quad \boldsymbol{\gamma_t} - \boldsymbol{\mu_\gamma} - \boldsymbol{K_4}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) = \boldsymbol{K_3}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta}) + \boldsymbol{\omega_t}, \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \\ &\Leftrightarrow \quad \boldsymbol{\gamma_t} - \boldsymbol{\mu_\gamma} - \boldsymbol{K_3}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta}) = \boldsymbol{K_4}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) + \boldsymbol{\omega_t}, \quad \boldsymbol{\omega_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \end{aligned}$$

which consequentially supports the acquisition of posterior results for $\boldsymbol{K_3}$ and $\boldsymbol{K_4}$ using standard multivariate regression results, respectively. More specific, we write $\boldsymbol{z_t} = \boldsymbol{\beta_t} - \boldsymbol{K_1}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta})$ and $\boldsymbol{x_t} = \boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}$ for $\boldsymbol{K_2}$, $\boldsymbol{z_t} = \boldsymbol{\gamma_t} - \boldsymbol{\mu_\gamma} - \boldsymbol{K_4}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma})$ and $\boldsymbol{x_t} = \boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta}$ for $\boldsymbol{K_3}$ and $\boldsymbol{z_t} = \boldsymbol{\gamma_t} - \boldsymbol{\mu_\gamma} - \boldsymbol{K_3}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta})$ and $\boldsymbol{x_t} = \boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}$ for $\boldsymbol{K_4}$. Herewith, all partitions fit the compact form framework

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{X} + \boldsymbol{U},$$

where $\boldsymbol{Z} = [\boldsymbol{z_1}, \boldsymbol{z_2}, \dots \boldsymbol{z_T}]$, $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_T}]$. We explicitly fill $\boldsymbol{U} = [\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \dots \boldsymbol{\eta_T}]$ for the sampling distribution of $\boldsymbol{K_2}$ and $\boldsymbol{U} = [\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \dots \boldsymbol{\omega_T}]$ for the posteriors of $\boldsymbol{K_3}$ and $\boldsymbol{K_4}$. Additionally, the dimensions of $\boldsymbol{Z}$ and $\boldsymbol{X}$ are updated from the completion of $\boldsymbol{z_t}$ and $\boldsymbol{x_t}$. The sampling distributions for all three partitions of the state dynamics matrix emerge from the procedures in Appendix D.2.2. We ensure the state equation to be covariance stationary via rejection sampling with each newly sampled partition, i.e. we construct $\boldsymbol{K}$ for each of the four partitions and seek resort to rejection sampling if a sampled partition disrupts the covariance stationarity of $\boldsymbol{K}$ (Diebold et al. (2006)).

F.2.2 Correlated Factor MF-RNS

Sampling of the state dynamics from the correlated factor MF-RNS model yields greater simplicity as we are no longer required to separately sample each partition of matrix $\boldsymbol{K}$. Hence, we rewrite the state equation as

$$\begin{aligned} \Leftrightarrow \quad & \boldsymbol{\xi_t} \quad = (\boldsymbol{I_{3+J}} - \boldsymbol{K})\boldsymbol{\mu} + \boldsymbol{K}\boldsymbol{\xi_{t-1}} + \boldsymbol{\epsilon_t} \\ \Leftrightarrow \quad & \boldsymbol{\xi_t} - \boldsymbol{\mu} = \boldsymbol{K}(\boldsymbol{\xi_{t-1}} - \boldsymbol{\mu}) + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}), \end{aligned}$$

where we stack both the latent factors and the observable series in the synthetic regressors $\boldsymbol{z_t} = \boldsymbol{\xi_t} - \boldsymbol{\mu}$ and $\boldsymbol{x_t} = \boldsymbol{\xi_{t-1}} - \boldsymbol{\mu}$. Forth, reformulation in compact form yields

$$\boldsymbol{Z} = \boldsymbol{K}\boldsymbol{X} + \boldsymbol{U}$$

with $\boldsymbol{Z} = [\boldsymbol{z_1}, \boldsymbol{z_2}, \dots \boldsymbol{z_T}]$, $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_T}]$ and $\boldsymbol{U} = [\boldsymbol{\eta_1}, \boldsymbol{\eta_2}, \dots \boldsymbol{\eta_T}]$ such that $\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{U} \in \mathbb{R}^{(J+P)\times T}$. Hence, we use the same vectorization techniques discussed in Appendix D.2.2 and assure the state equation to be covariance stationary via rejection of sampled matrices $\boldsymbol{K}$ with eigenvalues larger than one.

The block-diagonal structure of $\boldsymbol{W}$ supports independent sampling of its two components $\boldsymbol{Q}$ and $\boldsymbol{R}$. Hence, the full conditional marginal posterior distribution of a Cholesky-decomposable covariance matrix $\boldsymbol{Q}$ resolves to an $IW(\bar{\boldsymbol{Q}}, \bar{\nu}_Q)$ density with shape parameter

$$\bar{\boldsymbol{Q}} = \sum_{t=2}^{T} \left((\boldsymbol{\beta_t} - \boldsymbol{\mu_\beta}) - \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta} \\ \boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma} \end{bmatrix} \right) \left((\boldsymbol{\beta_t} - \boldsymbol{\mu_\beta}) - \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta} \\ \boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma} \end{bmatrix} \right)' + \underline{\boldsymbol{Q}}$$

and $\underline{\nu}_Q = (T-1) + \bar{\nu}_Q$ degrees of freedom in case we use an independent $IW(\underline{\boldsymbol{Q}}, \underline{\nu}_Q)$ prior.

F.3 State Sampling

The disturbance smoother from Durbin & Koopman (2012) limits the state equation to only comprise the VAR(1) representation of the latent factors. Hence, we write that

$$\begin{aligned} \boldsymbol{y_t} &= \boldsymbol{B}_{\boldsymbol{\lambda}}^{*}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon_t}, && \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \\ \boldsymbol{\beta_t} &= \boldsymbol{\mu_\beta} + \boldsymbol{K_1}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\mu_\beta}) + \boldsymbol{K_2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) + \boldsymbol{\eta_t}, \\ &= \boldsymbol{K_2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma}) + (\boldsymbol{I_3} - \boldsymbol{K_1})\boldsymbol{\mu_\beta} + \boldsymbol{K_1}\boldsymbol{\beta_{t-1}} + \boldsymbol{\eta_t}, && \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \qquad t = 1, \dots, T, \end{aligned}$$

which fits the framework from De Jong & Shephard (1995) and Durbin & Koopman (2002) via the additional inclusion of a time-varying constant with $\boldsymbol{K_2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\mu_\gamma})$. Hence, we refer to Appendix D for an explanation in greater detail.

G Posterior Sampling Scheme MF-AFRNS

The state space representation of the MF-AFRNS model corresponds to

$$\begin{pmatrix} \boldsymbol{y_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{0} \end{pmatrix} + \begin{bmatrix} \boldsymbol{B_\lambda^*} & \boldsymbol{0_{N\times P}} \\ \boldsymbol{0_{P\times 3}} & \boldsymbol{I_P} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon_t} \\ \boldsymbol{0_P} \end{pmatrix}, \quad \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}),$$
$$\begin{pmatrix} \boldsymbol{\beta_t} \\ \boldsymbol{\gamma_t} \end{pmatrix} = \left(\boldsymbol{I_{3+P}} - \begin{bmatrix} \boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \right) \begin{pmatrix} \boldsymbol{\theta_\beta} \\ \boldsymbol{\theta_\gamma} \end{pmatrix} + \begin{bmatrix} \boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta_{t-1}} \\ \boldsymbol{\gamma_{t-1}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta_t} \\ \boldsymbol{\omega_t} \end{pmatrix},$$

where $\boldsymbol{\gamma_t}$ $(P \times 1)$ bundles the macro-finance factors at time t, $\boldsymbol{B_\lambda^*}$ $(N \times 3)$ coincides with the rotated factor loadings, $\boldsymbol{H}$ $(N \times N)$ refers to a diagonal covariance matrix, matrices $\boldsymbol{e^{-K_1^{\mathbb{P}}\Delta t}}$ (3×3), $\boldsymbol{K_2}$ $(3 \times P)$, $\boldsymbol{K_3}$ $(P \times 3)$ and $\boldsymbol{K_4}$ $(P \times P)$ refer to the merged state dynamics between the unobserved factors and exogenous series, $\boldsymbol{\theta_\beta}$ (3×1) coincides with the unconditional means of the latent states, $\boldsymbol{\theta_\gamma}$ $(P \times 1)$ refers to the unconditional means of the exogenous factors, $\boldsymbol{Q}$ (3×3) denotes the diagonal covariance matrix on the upper part of the state equation and $\boldsymbol{R}$ $(P \times P)$ corresponds to a Cholesky-decomposable covariance matrix on the lower partition of the state equation. The assumed error structure for the state equation of the MF-AFRNS model is equivalent to the disturbance composition of the MF-RNS model in Appendix F. Our Bayesian approach only yields algebraic expressions for an independent factor variant of the MF-AFRNS model. Hence, we leave the implementation of a correlated factor MF-AFRNS model for future research.

G.1 Measurement Equation Sampling

The sampling of $\{a_i^2\}_{i=1}^N$ and $\{\sigma_i^2\}_{i=1}^N$ yields identical procedures for the AFRNS model and the MF-AFRNS model, respectively. Hence, we refer to Appendix E for an explanation in greater detail.

G.2 State Equation Sampling

Similar to Appendix E we subdivide the state equation of the independent factor MF-AFRNS model in its univariate components, i.e.

$$\begin{aligned} \Leftrightarrow \quad & \beta_{j,t} - \theta_{\beta,j} && = f_{jj}(\beta_{j,t} - \theta_{\beta,j}) + \boldsymbol{k_{j.}^2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\theta_\gamma}) + \eta_{j,t} \\ \Leftrightarrow \quad & \beta_{j,t} - \theta_{\beta,j} - \boldsymbol{k_{j.}^2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\theta_\gamma}) &&= f_{jj}(\beta_{j,t} - \theta_{\beta,j}) + \eta_{j,t}, \quad \eta_{j,t} \sim \mathcal{N}(0, q_{jj}^2), \quad j \in \{1,2,3\}, \end{aligned}$$

where $f_{jj} = e^{-\kappa_{jj}^{1,P}\Delta t}$ corresponds to a freely-sampled regression parameter, $\beta_{j,t}$ refers to the j-th factor at time t, $\theta_{\beta,j}$ coincides with the j-th element of $\boldsymbol{\theta_\beta}$ $\kappa_{jj}^{1,\mathbb{P}}$ denotes the j-th diagonal element of continuous-time state dynamics matrix $\boldsymbol{K_1^{\mathbb{P}}}$ and $\eta_{j,t}$ corresponds to the j-th element of the state disturbances with $\eta_{j,t} \sim \mathcal{N}(0, q_{jj}^2)$. From here, sampling of both f_{jj} and q_{jj}^2 resolves to standard univariate regression result with synthetically created variables $z_{j,t} = \beta_{j,t} - \theta_{\beta,j} - \boldsymbol{k_{j.}^2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\theta_\gamma})$ and $x_{j,t} = (\beta_{j,t} - \theta_{\beta,j})$ (e.g. see Appendix E).

The statistically limited structure of $\boldsymbol{Q}$ directly ties both the $\mathbb{Q}$-measure and the $\mathbb{P}$-measure via $\boldsymbol{\Sigma}$ and $\boldsymbol{K_1^{\mathbb{P}}}$, respectively. Hence, we extract the continuous-time state dynamics from $\kappa_{jj}^{1,\mathbb{P}} = -\log(f_{jj})/\Delta t \ \forall j$ and filter the risk-neutral volatilities via

$$\boldsymbol{Q} = \int_0^{\Delta t} \boldsymbol{e}^{-\boldsymbol{K_1^{\mathbb{P}}}s} \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \boldsymbol{e}^{-(\boldsymbol{K_1^{\mathbb{P}}})'s} \mathrm{d}s,$$

$$\begin{bmatrix} q_{11}^2 & 0 & 0 \\ 0 & q_{22}^2 & 0 \\ 0 & 0 & q_{33}^2 \end{bmatrix} = \int_0^{\Delta t} \begin{bmatrix} \sigma_{11}^2 e^{-\kappa_{11}^{1,\mathbb{P}} s} & 0 & 0 \\ 0 & \sigma_{22}^2 e^{-\kappa_{22}^{1,\mathbb{P}} s} & 0 \\ 0 & 0 & \sigma_{33}^2 e^{-\kappa_{33}^{1,\mathbb{P}} s} \end{bmatrix} \mathrm{d}s,$$

where we extract the j-th diagonal element of $\boldsymbol{\Sigma}$ by computing $\sigma_{jj} = \sqrt{\left(2q_{jj}^2 \kappa_{jj}^{1,\mathbb{P}}\right) / \left(1 - e^{-2\kappa_{jj}^{1,\mathbb{P}} \Delta t}\right)}$.

G.3 State Sampling

The state space representation for the MF-AFRNS model slightly changes compared to Appendix F due to the presence of the yield-adjustment, i.e.

$$\begin{aligned}
\boldsymbol{y_t} &= \boldsymbol{a} + \boldsymbol{B_\lambda^*}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon_t}, && \boldsymbol{\varepsilon_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \\
\boldsymbol{\beta_t} &= \boldsymbol{\theta_\beta} + \boldsymbol{e}^{-\boldsymbol{K_1^{\mathbb{P}}}\Delta t}(\boldsymbol{\beta_{t-1}} - \boldsymbol{\theta_\beta}) + \boldsymbol{K_2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\theta_\gamma}) + \boldsymbol{\eta_t}, \\
&= \boldsymbol{K_2}(\boldsymbol{\gamma_{t-1}} - \boldsymbol{\theta_\gamma}) + (\boldsymbol{I_3} - \boldsymbol{e}^{-\boldsymbol{K_1^{\mathbb{P}}}\Delta t})\boldsymbol{\theta_\beta} + \boldsymbol{e}^{-\boldsymbol{K_1^{\mathbb{P}}}\Delta t}\boldsymbol{\beta_{t-1}} + \boldsymbol{\eta_t}, && \boldsymbol{\eta_t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \quad t = 1, \dots, T.
\end{aligned}$$

where we use these equations to run the simulation smoother from Durbin & Koopman (2002). Hence, we refer to the state smoothing procedures in Appendix D for an explanation in greater detail.

H Tables & Figures

H.1 Tables

Table 5: Parameter estimates of the RNS model

	Rotated Nelson-Siegel (RNS)													
Statistic	**Independent factors**							**Correlated factors**						
	$\boldsymbol{F_{1,1}}$	$\boldsymbol{F_{1,2}}$	$\boldsymbol{F_{1,3}}$	$\boldsymbol{\mu_1}$	$\boldsymbol{Q_{1,1}}$	$\boldsymbol{Q_{1,2}}$	$\boldsymbol{Q_{1,3}}$	$\boldsymbol{F_{1,1}}$	$\boldsymbol{F_{1,2}}$	$\boldsymbol{F_{1,3}}$	$\boldsymbol{\mu_1}$	$\boldsymbol{Q_{1,1}}$	$\boldsymbol{Q_{1,2}}$	$\boldsymbol{Q_{1,3}}$
mean	0.992	0	0	0.015	0.328	0	0	0.995	0.027	0.049	0.014	0.246	-0.244	-0.043
std.	0.005	-	-	0.008	0.037	-	-	0.010	0.010	0.007	0.007	0.028	0.065	0.137
$HPDI_{lo}$	0.982	-	-	-0.002	0.257	-	-	0.975	0.008	0.037	-0.001	0.192	-0.368	-0.310
$HPDI_{up}$	1.000	-	-	0.031	0.399	-	-	1.015	0.045	0.063	0.026	0.302	-0.114	0.227
Geweke CD	-0.111	-	-	0.197	0.869	-	-	-0.210	1.015	0.395	0.452	0.585	-0.081	0.411
	$\boldsymbol{F_{2,1}}$	$\boldsymbol{F_{2,2}}$	$\boldsymbol{F_{2,3}}$	$\boldsymbol{\mu_2}$	$\boldsymbol{Q_{2,1}}$	$\boldsymbol{Q_{2,2}}$	$\boldsymbol{Q_{2,3}}$	$\boldsymbol{F_{2,1}}$	$\boldsymbol{F_{2,2}}$	$\boldsymbol{F_{2,3}}$	$\boldsymbol{\mu_2}$	$\boldsymbol{Q_{2,1}}$	$\boldsymbol{Q_{2,2}}$	$\boldsymbol{Q_{2,3}}$
mean	0	0.974	0	0.030	0	2.487	0	-0.009	0.947	-0.032	0.030	-0.244	2.505	-3.118
std.	-	0.013	-	0.008	-	0.255	-	0.031	0.029	0.021	0.007	0.065	0.257	0.445
$HPDI_{lo}$	-	0.949	-	0.013	-	1.999	-	-0.067	0.889	-0.072	0.017	-0.368	2.020	-3.977
$HPDI_{up}$	-	0.998	-	0.046	-	2.994	-	0.052	1.003	0.009	0.044	-0.114	3.009	-2.256
Geweke CD	-	-0.192	-	-0.945	-	-0.071	-	0.532	0.108	0.427	0.300	-0.081	-0.603	-0.382
	$\boldsymbol{F_{3,1}}$	$\boldsymbol{F_{3,2}}$	$\boldsymbol{F_{3,3}}$	$\boldsymbol{\mu_3}$	$\boldsymbol{Q_{3,1}}$	$\boldsymbol{Q_{3,2}}$	$\boldsymbol{Q_{3,3}}$	$\boldsymbol{F_{3,1}}$	$\boldsymbol{F_{3,2}}$	$\boldsymbol{F_{3,3}}$	$\boldsymbol{\mu_3}$	$\boldsymbol{Q_{3,1}}$	$\boldsymbol{Q_{3,2}}$	$\boldsymbol{Q_{3,3}}$
mean	0	0	0.926	-0.028	0	0	9.539	-0.049	-0.046	0.900	-0.029	-0.043	-3.118	10.231
std.	-	-	0.027	0.007	-	-	1.054	0.061	0.058	0.041	0.004	0.137	0.445	1.158
$HPDI_{lo}$	-	-	0.871	-0.042	-	-	7.522	-0.164	-0.165	0.821	-0.038	-0.310	-3.977	7.958
$HPDI_{up}$	-	-	0.978	-0.014	-	-	11.615	0.075	0.062	0.981	-0.020	0.227	-2.256	12.436
Geweke CD	-	-	1.034	-0.901	-	-	0.747	-0.685	-0.434	-0.692	-0.263	0.411	-0.382	-0.389

Note: This table contains the acquired posterior means, standard deviations, HPDIs and MCMC convergence diagnostics (e.g. see Geweke et al. (1992)) for both RNS models. We plug in $\lambda = 0.3844$ from the initial two-step routine for maturities in years (Diebold & Li (2006); Çakmaklı (2020)). The reported estimates are acquired via the procedures in Sections 3 and 5.1, respectively. We multiply the reported point estimates, standard deviations and HPDIs from covariance matrix $\boldsymbol{Q}$ by 10^5 for notational convenience.

Table 6: Parameter estimates of the AFRNS model

	Arbitrage-Free Rotated Nelson-Siegel (AFRNS)							
Statistic	**State dynamics**		**$A_0(3)$ parameters**			**Yield-adjustment**		
	$e^{-\kappa^{\mathbb{P}}_{11}\Delta t}$	$Q_{1,1}$	$k^{\mathbb{P}}_{11}$	$\Sigma_{1,1}$	$\theta^{\mathbb{P}}_{1}$	a_{3M}	a_{6M}	a_{1Y}
mean	0.991	0.379	0.107	0.007	0.015	-0.005	-0.023	-0.123
std.	0.006	0.039	0.071	0.000	0.008	0.000	0.002	0.009
HPD_{lo}	0.980	0.307	0.000	0.006	-0.003	-0.006	-0.027	-0.140
$HPDI_{up}$	1.000	0.457	0.241	0.007	0.031	-0.004	-0.019	-0.105
Geweke CD	-0.291	1.936	0.293	1.986	-0.012	-1.972	-2.031	-2.048
	$e^{-\kappa^{\mathbb{P}}_{22}\Delta t}$	$Q_{2,2}$	$k^{\mathbb{P}}_{22}$	$\Sigma_{2,2}$	$\theta^{\mathbb{P}}_{1}$	a_{2Y}	a_{3Y}	a_{5Y}
mean	0.975	2.147	0.303	0.016	0.030	-0.809	-2.467	-9.165
std.	0.013	0.216	0.156	0.001	0.008	0.053	0.162	0.589
HPD_{lo}	0.952	1.753	0.001	0.015	0.014	-0.913	-2.788	-10.303
$HPDI_{up}$	0.999	2.588	0.577	0.018	0.046	-0.707	-2.159	-8.017
Geweke CD	-0.205	-0.961	0.195	-0.973	-0.643	-1.654	-1.325	-0.949
	$e^{-\kappa^{\mathbb{P}}_{33}\Delta t}$	$Q_{3,3}$	$k^{\mathbb{P}}_{33}$	$\Sigma_{3,3}$	$\theta^{\mathbb{P}}_{1}$	a_{7Y}	a_{10Y}	
mean	0.925	10.216	0.940	0.036	-0.033	-20.012	-42.969	
std.	0.029	1.049	0.382	0.002	0.008	1.258	2.712	
HPD_{lo}	0.870	8.307	0.194	0.033	-0.050	-22.552	-48.509	
$HPDI_{up}$	0.984	12.381	1.677	0.040	-0.018	-17.667	-37.883	
Geweke CD	0.230	1.022	-0.225	0.925	1.242	-0.673	-0.320	

Note: This table contains the acquired posterior means, standard deviations, HPDIs and MCMC convergence diagnostics (e.g. see Geweke et al. (1992)) for the AFRNS model. We plug in $\lambda = 0.3844$ from the initial two-step routine for maturities in years (Diebold & Li (2006); Çakmaklı (2020)). The reported estimates are acquired via the procedures in Sections 3 and 5.1, respectively. We multiply the reported point estimates, standard deviations and HPDIs from covariance matrix $\boldsymbol{Q}$ by 10^5 for notational convenience. In a similar fashion, we report the point estimates from the yield-adjustment $\boldsymbol{a}$ in basis points.

Table 7: Parameter estimates of the MF-RNS model

	Macro-Finance Rotated Nelson-Siegel (MF-RNS)																			
Statistic	**Independent factors**										**Correlated factors**									
	$\boldsymbol{K_{1,1}}$	$\boldsymbol{K_{1,2}}$	$\boldsymbol{K_{1,3}}$	$\boldsymbol{K_{1,4}}$	$\boldsymbol{K_{1,5}}$	$\boldsymbol{K_{1,6}}$	$\boldsymbol{\mu_1}$	$\boldsymbol{Q_{1,1}}$	$\boldsymbol{Q_{1,2}}$	$\boldsymbol{Q_{1,3}}$	$\boldsymbol{K_{1,1}}$	$\boldsymbol{K_{1,2}}$	$\boldsymbol{K_{1,3}}$	$\boldsymbol{K_{1,4}}$	$\boldsymbol{K_{1,5}}$	$\boldsymbol{K_{1,6}}$	$\boldsymbol{\mu_1}$	$\boldsymbol{Q_{1,1}}$	$\boldsymbol{Q_{1,2}}$	$\boldsymbol{Q_{1,3}}$
mean	0.978	0	0	-0.031	0.036	0.004	0.017	0.307	0	0	0.956	0.008	0.046	-0.038	0.027	0.008	0.017	0.223	-0.212	-0.052
std.	0.013	-	-	0.040	0.010	0.002	0.005	0.034	-	-	0.017	0.013	0.006	0.038	0.009	0.002	0.004	0.027	0.062	0.136
$HPDI_{lo}$	0.953	-	-	-0.105	0.018	0.000	0.007	0.244	-	-	0.924	-0.019	0.033	-0.109	0.010	0.004	0.009	0.174	-0.334	-0.336
$HPDI_{up}$	1.000	-	-	0.051	0.056	0.009	0.026	0.376	-	-	0.989	0.032	0.058	0.037	0.044	0.012	0.024	0.276	-0.091	0.198
Geweke CD	-0.355	-	-	-1.783	1.452	-0.332	-0.040	1.152	-	-	-0.748	-0.525	-0.410	-0.899	2.293	2.043	0.416	-1.161	0.538	-0.860
	$\boldsymbol{K_{2,1}}$	$\boldsymbol{K_{2,2}}$	$\boldsymbol{K_{2,3}}$	$\boldsymbol{K_{2,4}}$	$\boldsymbol{K_{2,5}}$	$\boldsymbol{K_{2,6}}$	$\boldsymbol{\mu_2}$	$\boldsymbol{Q_{2,1}}$	$\boldsymbol{Q_{2,2}}$	$\boldsymbol{Q_{2,3}}$	$\boldsymbol{K_{2,1}}$	$\boldsymbol{K_{2,2}}$	$\boldsymbol{K_{2,3}}$	$\boldsymbol{K_{2,4}}$	$\boldsymbol{K_{2,5}}$	$\boldsymbol{K_{2,6}}$	$\boldsymbol{\mu_2}$	$\boldsymbol{Q_{2,1}}$	$\boldsymbol{Q_{2,2}}$	$\boldsymbol{Q_{2,3}}$
mean	0	0.939	0	-0.066	-0.074	-0.005	0.032	0	2.434	0	0.015	0.913	-0.035	-0.127	-0.077	-0.007	0.031	-0.212	2.444	-3.127
std.	-	0.024	-	0.114	0.029	0.004	0.006	-	0.246	-	0.055	0.043	0.021	0.121	0.029	0.007	0.006	0.062	0.249	0.445
$HPDI_{lo}$	-	0.893	-	-0.291	-0.132	-0.013	0.020	-	1.977	-	-0.095	0.830	-0.075	-0.360	-0.134	-0.021	0.019	-0.334	1.976	-4.022
$HPDI_{up}$	-	0.985	-	0.158	-0.020	0.003	0.045	-	2.925	-	0.120	0.996	0.006	0.116	-0.023	0.007	0.042	-0.091	2.939	-2.304
Geweke CD	-	0.135	-	-0.790	-0.811	-0.587	0.257	-	0.832	-	-0.324	-0.093	0.528	-0.016	-0.638	0.378	-0.752	0.538	-1.044	0.610
	$\boldsymbol{K_{3,1}}$	$\boldsymbol{K_{3,2}}$	$\boldsymbol{K_{3,3}}$	$\boldsymbol{K_{3,4}}$	$\boldsymbol{K_{3,5}}$	$\boldsymbol{K_{3,6}}$	$\boldsymbol{\mu_3}$	$\boldsymbol{Q_{3,1}}$	$\boldsymbol{Q_{3,2}}$	$\boldsymbol{Q_{3,3}}$	$\boldsymbol{K_{3,1}}$	$\boldsymbol{K_{3,2}}$	$\boldsymbol{K_{3,3}}$	$\boldsymbol{K_{3,4}}$	$\boldsymbol{K_{3,5}}$	$\boldsymbol{K_{3,6}}$	$\boldsymbol{\mu_3}$	$\boldsymbol{Q_{3,1}}$	$\boldsymbol{Q_{3,2}}$	$\boldsymbol{Q_{3,3}}$
mean	0	0	0.909	0.070	0.026	-0.008	-0.026	0	0	9.623	-0.032	-0.005	0.904	0.111	0.034	-0.005	-0.027	-0.052	-3.127	10.402
std.	-	-	0.032	0.180	0.050	0.008	0.006	-	-	1.078	0.108	0.085	0.042	0.246	0.058	0.014	0.004	0.136	0.445	1.192
$HPDI_{lo}$	-	-	0.847	-0.286	-0.073	-0.023	-0.038	-	-	7.646	-0.245	-0.171	0.820	-0.364	-0.075	-0.033	-0.034	-0.336	-4.022	8.135
$HPDI_{up}$	-	-	0.971	0.423	0.123	0.009	-0.014	-	-	11.788	0.180	0.158	0.982	0.605	0.152	0.022	-0.019	0.198	-2.304	12.708
Geweke CD	-	-	0.232	1.641	-0.862	-0.263	-0.459	-	-	-0.083	-0.141	-0.704	-0.328	0.363	0.604	-0.520	0.912	-0.860	0.610	0.533
	$\boldsymbol{K_{4,1}}$	$\boldsymbol{K_{4,2}}$	$\boldsymbol{K_{4,3}}$	$\boldsymbol{K_{4,4}}$	$\boldsymbol{K_{4,5}}$	$\boldsymbol{K_{4,6}}$	$\boldsymbol{\mu_4}$	$\boldsymbol{R_{1,1}}$	$\boldsymbol{R_{1,2}}$	$\boldsymbol{R_{1,3}}$	$\boldsymbol{K_{4,1}}$	$\boldsymbol{K_{4,2}}$	$\boldsymbol{K_{4,3}}$	$\boldsymbol{K_{4,4}}$	$\boldsymbol{K_{4,5}}$	$\boldsymbol{K_{4,6}}$	$\boldsymbol{\mu_4}$	$\boldsymbol{R_{1,1}}$	$\boldsymbol{R_{1,2}}$	$\boldsymbol{R_{1,3}}$
mean	0.000	-0.002	0.001	0.934	0.018	0.002	0.020	0.116	-0.018	0.018	-0.001	-0.003	0.001	0.934	0.017	0.002	0.020	0.116	-0.018	0.019
std.	0.011	0.009	0.004	0.025	0.006	0.001	0.001	0.012	0.014	0.041	0.012	0.009	0.005	0.026	0.006	0.002	0.001	0.012	0.014	0.041
$HPDI_{lo}$	-0.021	-0.019	-0.007	0.884	0.006	-0.001	0.018	0.094	-0.045	-0.064	-0.023	-0.020	-0.008	0.881	0.006	-0.001	0.018	0.094	-0.045	-0.065
$HPDI_{up}$	0.022	0.015	0.010	0.980	0.030	0.004	0.022	0.140	0.009	0.098	0.023	0.016	0.010	0.983	0.030	0.005	0.021	0.141	0.009	0.095
Geweke CD	1.545	0.945	-0.709	0.716	-0.211	-0.979	0.006	0.691	0.823	0.239	-0.482	0.796	0.200	-0.144	0.789	-0.311	1.217	-0.159	1.012	-0.4001
	$\boldsymbol{K_{5,1}}$	$\boldsymbol{K_{5,2}}$	$\boldsymbol{K_{5,3}}$	$\boldsymbol{K_{5,4}}$	$\boldsymbol{K_{5,5}}$	$\boldsymbol{K_{5,6}}$	$\boldsymbol{\mu_5}$	$\boldsymbol{R_{2,1}}$	$\boldsymbol{R_{2,2}}$	$\boldsymbol{R_{2,3}}$	$\boldsymbol{K_{5,1}}$	$\boldsymbol{K_{5,2}}$	$\boldsymbol{K_{5,3}}$	$\boldsymbol{K_{5,4}}$	$\boldsymbol{K_{5,5}}$	$\boldsymbol{K_{5,6}}$	$\boldsymbol{\mu_5}$	$\boldsymbol{R_{2,1}}$	$\boldsymbol{R_{2,2}}$	$\boldsymbol{R_{2,3}}$
mean	-0.022	-0.011	0.017	-0.207	1.003	0.004	0.008	-0.018	0.298	-0.049	-0.023	-0.011	0.017	-0.204	1.004	0.004	0.007	-0.018	0.298	-0.049
std.	0.019	0.015	0.007	0.043	0.010	0.002	0.004	0.014	0.031	0.066	0.019	0.015	0.007	0.043	0.010	0.003	0.004	0.014	0.031	0.066
$HPDI_{lo}$	-0.059	-0.038	0.002	-0.294	0.984	-0.001	0.001	-0.045	0.241	-0.182	-0.062	-0.039	0.002	-0.290	0.984	-0.001	-0.003	-0.045	0.236	-0.179
$HPDI_{up}$	0.014	0.019	0.030	-0.126	1.023	0.009	0.016	0.009	0.361	0.078	0.013	0.018	0.031	-0.121	1.023	0.009	0.015	0.009	0.355	0.080
Geweke CD	-0.828	-0.113	1.980	0.835	0.104	1.486	-0.264	0.823	0.554	-1.335	-0.822	-1.454	-1.680	-0.894	-0.639	-1.224	-0.443	1.012	-1.756	-1.486
	$\boldsymbol{K_{6,1}}$	$\boldsymbol{K_{6,2}}$	$\boldsymbol{K_{6,3}}$	$\boldsymbol{K_{6,4}}$	$\boldsymbol{K_{6,5}}$	$\boldsymbol{K_{6,6}}$	$\boldsymbol{\mu_6}$	$\boldsymbol{R_{3,1}}$	$\boldsymbol{R_{3,2}}$	$\boldsymbol{R_{3,3}}$	$\boldsymbol{K_{6,1}}$	$\boldsymbol{K_{6,2}}$	$\boldsymbol{K_{6,3}}$	$\boldsymbol{K_{6,4}}$	$\boldsymbol{K_{6,5}}$	$\boldsymbol{K_{6,6}}$	$\boldsymbol{\mu_6}$	$\boldsymbol{R_{3,1}}$	$\boldsymbol{R_{3,2}}$	$\boldsymbol{R_{3,3}}$
mean	0.147	0.082	0.095	-0.202	-0.063	0.979	0.766	0.018	-0.049	2.671	0.154	0.084	0.093	-0.209	-0.063	0.979	0.766	0.019	-0.048	2.674
std.	0.058	0.045	0.022	0.127	0.031	0.007	0.010	0.041	0.066	0.278	0.058	0.045	0.022	0.127	0.031	0.007	0.010	0.041	0.066	0.276
$HPDI_{lo}$	0.036	-0.004	0.053	-0.456	-0.125	0.965	0.747	-0.064	-0.182	2.170	0.037	-0.007	0.050	-0.466	-0.121	0.964	0.746	-0.065	-0.179	2.155
$HPDI_{up}$	0.262	0.173	0.139	0.040	-0.003	0.994	0.785	0.098	0.078	3.245	0.265	0.171	0.137	0.028	-0.002	0.993	0.784	0.095	0.079	3.223
Geweke CD	0.526	0.016	-0.933	-0.094	-0.860	-0.422	-1.376	0.239	-1.335	-0.222	-0.067	-0.509	0.579	-0.550	0.144	0.210	2.293	-0.401	-1.486	-0.593

Note: This table contains the acquired posterior means, standard deviations, HPDIs and MCMC convergence diagnostics (e.g. see Geweke et al. (1992)) for both MF-RNS models. We plug in $\lambda = 0.3844$ from the initial two-step routine for maturities in years (Diebold & Li (2006); Çakmaklı (2020)). The reported estimates are acquired via the procedures in Sections 3 and 5.1, respectively. The factor order corresponds to [Short Rate, Slope, Curvature, CCPI, NFP, MF] (e.g. see Section 4). We multiply the reported point estimates, standard deviations and HPDIs from covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ by 10^5 for notational convenience.

Table 8: Parameter estimates of the MF-AFRNS model

	Macro-Finance Arbitrage-Free Rotated Nelson-Siegel (MF-AFRNS)															
Statistic	**State dynamics**									**$A_0(3)$ parameters**				**Yield-adjustment**		
	$e^{-\kappa_{11}^{1,\mathbb{P}}\Delta t}$	$K_{1,2}$	$K_{1,3}$	$K_{1,4}$	$K_{1,5}$	$K_{1,6}$	$Q_{1,1}$	$Q_{1,2}$	$Q_{1,3}$	$K_{1,1}^{\mathbb{P}}$	$K_{1,2}^{\mathbb{P}}$	$K_{1,3}^{\mathbb{P}}$	$\theta_1^{\mathbb{P}}$	a_{3M}	a_{6M}	a_{1Y}
mean	0.974	0	0	-0.022	0.038	0.005	0.346	0	0	0.312	0	0	0.018	-0.005	-0.021	-0.118
std.	0.013	-	-	0.041	0.010	0.002	0.037	-	-	0.162	-	-	0.005	0.000	0.002	0.009
$HPDI_{lo}$	0.948	-	-	-0.099	0.018	0.000	0.278	-	-	0.015	-	-	0.009	-0.006	-0.025	-0.135
$HPDI_{up}$	0.999	-	-	0.060	0.059	0.009	0.421	-	-	0.647	-	-	0.027	-0.004	-0.018	-0.099
Geweke CD	0.387	-	-	0.984	-1.245	-0.331	-3.070	-		-0.391	-	-	1.190	3.078	3.047	2.764
	$K_{2,1}$	$e^{-\kappa_{22}^{1,\mathbb{P}}\Delta t}$	$K_{2,3}$	$K_{2,4}$	$K_{2,5}$	$K_{2,6}$	$Q_{2,1}$	$Q_{2,2}$	$Q_{2,3}$	$K_{2,1}^{\mathbb{P}}$	$K_{2,2}^{\mathbb{P}}$	$K_{2,3}^{\mathbb{P}}$	$\theta_2^{\mathbb{P}}$	a_{2Y}	a_{3Y}	a_{5Y}
mean	0	0.938	0	-0.069	-0.074	-0.005	0	2.435	0	0	0.767	0	0.035	-0.805	-2.530	-9.797
std.	-	0.024	-	0.113	0.029	0.004	-	0.249	-	-	0.302	-	0.006	0.057	0.181	0.693
$HPDI_{lo}$	-	0.890	-	-0.289	-0.129	-0.014	-	1.991	-	-	0.211	-	0.022	-0.922	-2.906	-11.144
$HPDI_{up}$	-	0.983	-	0.152	-0.018	0.003	-	2.938	-	-	1.396	-	0.047	-0.698	-2.196	-8.435
Geweke CD	-	0.077	-	0.934	0.603	0.250	-	-0.905	-	-	-0.067	-	-0.220	1.956	1.560	1.360
	$K_{3,1}$	$K_{3,2}$	$e^{-\kappa_{33}^{1,\mathbb{P}}\Delta t}$	$K_{3,4}$	$K_{3,5}$	$K_{3,6}$	$Q_{3,1}$	$Q_{3,2}$	$Q_{3,3}$	$K_{3,1}^{\mathbb{P}}$	$K_{3,2}^{\mathbb{P}}$	$K_{3,3}^{\mathbb{P}}$	$\theta_3^{\mathbb{P}}$	a_{7Y}	a_{10Y}	
mean	0	0	0.904	0.107	0.030	-0.007	-	-	11.542	0	0	1.213	-0.033	-21.930	-47.967	
std.	-	-	0.034	0.194	0.056	0.009	-	-	1.212	-	-	0.450	0.007	1.514	3.291	
$HPDI_{lo}$	-	-	0.838	-0.297	-0.078	-0.025	-	-	9.247	-	-	0.380	-0.046	-24.925	-54.621	
$HPDI_{up}$	-	-	0.969	0.464	0.140	0.010	-	-	13.936	-	-	2.121	-0.019	-19.013	-41.790	
Geweke CD	-	-	0.018	-0.487	1.373	1.310	-	-	-0.137	-	-	-0.042	0.816	1.380	1.463	
	$K_{4,1}$	$K_{4,2}$	$K_{4,3}$	$K_{4,4}$	$K_{4,5}$	$K_{4,6}$	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{1,3}$	$\theta_4^{\mathbb{P}}$			
mean	0.000	-0.002	0.001	0.934	0.018	0.002	0.116	-0.018	0.020	0.007	0	0	0.020			
std.	0.011	0.009	0.004	0.025	0.006	0.001	0.012	0.014	0.041	0.000	-	-	0.001			
$HPDI_{lo}$	-0.022	-0.020	-0.007	0.885	0.006	-0.001	0.094	-0.045	-0.060	0.006	-	-	0.018			
$HPDI_{up}$	0.022	0.015	0.009	0.982	0.030	0.005	0.140	0.009	0.102	0.007	-	-	0.023			
Geweke CD	-0.772	-1.339	-0.947	-0.537	-0.404	0.847	0.628	-0.493	-0.363	-3.126	-	-	-0.539			
	$K_{5,1}$	$K_{5,2}$	$K_{5,3}$	$K_{5,4}$	$K_{5,5}$	$K_{5,6}$	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$	$\Sigma_{2,1}$	$\Sigma_{2,2}$	$\Sigma_{2,3}$	$\theta_5^{\mathbb{P}}$			
mean	-0.022	-0.010	0.017	-0.204	1.003	0.004	-0.018	0.295	-0.046	0	0.018	0	0.010			
std.	0.019	0.015	0.007	0.042	0.010	0.002	0.014	0.031	0.066	-	0.001	-	0.004			
$HPDI_{lo}$	-0.061	-0.037	0.004	-0.287	0.983	-0.001	-0.045	0.240	-0.176	-	0.016	-	0.002			
$HPDI_{up}$	0.013	0.020	0.032	-0.122	1.022	0.009	0.009	0.358	0.082	-	0.020	-	0.018			
Geweke CD	-2.042	-0.470	0.486	0.802	1.084	2.684	-0.493	0.988	2.958	-	-0.904	-	-0.071			
	$K_{6,1}$	$K_{6,2}$	$K_{6,3}$	$K_{6,4}$	$K_{6,5}$	$K_{6,6}$	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$	$\Sigma_{3,1}$	$\Sigma_{3,2}$	$\Sigma_{3,3}$	$\theta_6^{\mathbb{P}}$			
mean	0.153	0.080	0.089	-0.215	-0.064	0.978	0.020	-0.046	2.684	0	0	0.039	0.765			
std.	0.059	0.046	0.021	0.128	0.031	0.007	0.041	0.066	0.279	-	-	0.002	0.010			
$HPDI_{lo}$	0.036	-0.014	0.047	-0.472	-0.126	0.964	-0.060	-0.176	2.147	-	-	0.035	0.746			
$HPDI_{up}$	0.265	0.163	0.130	0.031	-0.005	0.993	0.102	0.082	3.231	-	-	0.043	0.785			
Geweke CD	0.561	0.155	-1.375	0.135	0.814	0.060	-0.363	2.958	-0.787	-	-	-0.163	-0.818			

Note: This table contains the acquired posterior means, standard deviations, HPDIs and MCMC convergence diagnostics (e.g. see Geweke et al. (1992)) for the MF-AFRNS model. We plug in $\lambda = 0.3844$ from the initial two-step routine for maturities in years (Diebold & Li (2006); Çakmaklı (2020)). The reported estimates are acquired via the procedures in Sections 3 and 5.1, respectively. The factor order corresponds to [Short Rate, Slope, Curvature, CCPI, NFP, MF] (e.g. see Section 4). We multiply the reported point estimates, standard deviations and HPDIs from covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ by 10^5 for notational convenience. In a similar fashion, we report the point estimates from the yield-adjustment $\boldsymbol{a}$ in basis points.

Table 9: In-sample fit from Maximum Likelihood

	Rotated Nelson-Siegel (RNS)						**Dynamic Nelson-Siegel (DNS)**					
	Independent		**Correlated**		**Arbitrage-Free**		**Independent**		**Correlated**		**Arbitrage-Free**	
Maturity	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
3	-14.57	23.77	-14.58	23.81	-15.00	23.89	-15.13	24.28	-14.59	23.82	-14.20	23.08
6	-4.63	10.25	-4.69	10.28	-4.93	10.22	-4.95	10.10	-4.69	10.28	-4.30	9.93
12	0.58	1.55	0.48	1.21	0.50	1.84	0.53	1.45	0.47	1.20	0.85	2.64
24	0.12	2.01	0.05	2.05	0.25	2.07	0.21	1.96	0.05	2.05	0.29	1.89
36	-0.00	0.00	-0.00	0.00	0.00	0.04	-0.00	0.00	-0.00	0.00	-0.01	0.39
60	-0.25	2.36	-0.18	2.39	-0.73	2.58	-0.39	2.35	-0.17	2.39	-0.60	2.59
84	0.00	0.00	0.00	0.00	-0.01	0.12	0.00	0.00	0.00	0.00	-0.07	2.29
120	0.88	8.14	0.62	8.24	5.62	10.54	1.44	8.05	0.62	8.24	3.43	9.19
All	-2.23	6.01	-2.29	6.01	-1.79	6.40	-2.29	6.08	-2.29	6.00	-1.83	6.50

Note: This table contains the mean residuals (Mean) and Root Mean Squared Errors (RMSEs) in basis points (bp) over the period July 2003 to February 2020 for all yields-only specifications and maturities considered. The in-sample fits reported in this table follow from classical estimation (e.g. see Appendix C.3). The mean residuals are constructed as the difference between the realized yields minus the fitted yield. Italic (Bold) entries correspond to the cross-sectional absolute lowest mean residual (RMSE) for that maturity.

Table 10: Root Mean Squared Forecast Errors (RMSFEs)

	Maturity (Months)								
Model	**3**	**6**	**12**	**24**	**36**	**60**	**84**	**120**	**All**
Panel A: One-month-ahead forecasts									
RNS_{indep}	**14.09**	14.86	18.30	19.48	20.62	21.71	21.65	21.87	19.07
RNS_{corr}	17.02	13.79	14.69	18.00	19.94	21.43	21.63	22.73	18.65
AFRNS	18.87	14.66	16.95	20.18	21.74	23.06	23.13	26.34	20.62
DNS_{indep}	14.13	14.92	18.35	19.54	20.68	21.79	21.71	21.87	19.13
DNS_{corr}	16.99	13.73	**14.61**	17.97	19.93	21.43	21.67	22.84	18.65
AFNS	15.34	**13.00**	15.36	**17.87**	**19.32**	**20.66**	**20.75**	**20.86**	**17.90**
Panel B: One-quarter-ahead forecasts									
RNS_{indep}	30.84	35.04	39.65	40.84	41.57	41.29	39.87	38.48	38.45
RNS_{corr}	29.71	29.66	32.63	37.36	39.37	39.64	38.84	39.47	35.83
AFRNS	34.31	36.07	41.15	45.19	46.91	47.64	47.26	47.82	43.29
DNS_{indep}	30.85	35.12	39.73	40.89	41.59	41.23	39.74	38.27	38.43
DNS_{corr}	29.54	29.52	32.47	37.17	39.16	39.45	38.78	39.68	35.72
AFNS	**23.78**	**25.59**	**29.27**	**31.91**	**33.94**	**35.73**	**35.80**	**35.15**	**31.40**
Panel C: Six-months-ahead forecasts									
RNS_{indep}	57.63	63.53	68.83	69.56	68.78	64.77	60.03	55.17	63.54
RNS_{corr}	54.83	57.36	61.99	67.04	67.98	65.01	61.55	60.21	62.00
AFRNS	67.49	71.76	77.61	81.58	82.49	81.20	79.17	76.87	77.27
DNS_{indep}	57.61	63.45	68.72	69.44	68.65	64.64	59.89	55.10	63.44
DNS_{corr}	54.56	57.13	61.75	66.72	67.59	64.62	61.35	60.39	61.77
AFNS	**42.57**	**46.94**	**51.41**	**53.76**	**55.12**	**55.25**	**53.68**	**50.68**	**51.18**
Panel D: One-year-ahead forecasts									
RNS_{indep}	96.71	104.45	110.38	108.39	103.48	91.30	80.39	70.35	95.68
RNS_{corr}	106.14	110.10	113.54	115.89	113.04	103.14	95.17	91.16	106.02
AFRNS	118.81	125.38	131.24	132.87	130.74	124.50	119.57	114.79	124.74
DNS_{indep}	96.69	104.35	110.30	108.25	103.30	91.06	80.10	70.03	95.51
DNS_{corr}	105.72	109.76	113.12	115.19	112.12	102.02	94.17	90.77	105.36
AFNS	**77.02**	**83.68**	**88.94**	**89.36**	**87.34**	**81.25**	**75.26**	**67.94**	**81.35**

Note: This table contains the Root Mean Squared Forecast Errors (RMSFEs) in basis points across maturities for all specifications and forecast horizons considered. The maturities are reported in months. The out-of-sample period encapsulates the period between January 2016 and February 2020. All specifications and forecasts are estimated according to Section 3. Bold entries indicate the cross-sectional lowest RMSFE for a respective horizon and maturity.

Table 11: Root Mean Squared Forecast Errors (RMSFEs)

	Maturity (Months)								
Model	**3**	**6**	**12**	**24**	**36**	**60**	**84**	**120**	**All**
Panel A: One-month-ahead forecasts									
MF-RNS$_{indep}$	**16.08**	15.40	17.82	19.67	20.87	21.86	22.14	23.34	19.65
MF-RNS$_{corr}$	17.56	13.14	**13.25**	**17.08**	18.99	20.62	21.62	24.31	18.32
MF-AFRNS	19.57	15.28	16.63	19.65	21.04	22.57	22.64	23.44	20.10
MF-DNS$_{indep}$	17.62	16.63	17.96	19.73	20.70	21.43	21.74	**23.23**	19.88
MF-DNS$_{corr}$	17.28	**13.10**	13.45	17.09	**18.93**	**20.51**	**21.52**	24.18	**18.26**
MF-AFNS	20.48	16.55	17.21	19.56	20.44	21.44	21.47	24.45	20.20
Panel B: One-quarter-ahead forecasts									
MF-RNS$_{indep}$	34.41	36.88	40.26	41.89	42.39	42.09	41.98	43.53	40.43
MF-RNS$_{corr}$	**26.39**	**25.87**	**28.38**	**32.80**	34.55	35.75	**37.57**	42.51	**32.98**
MF-AFRNS	33.52	34.84	38.67	41.86	42.98	43.90	44.03	41.62	40.18
MF-DNS$_{indep}$	34.20	36.11	37.96	39.07	39.17	38.50	38.39	40.30	37.96
MF-DNS$_{corr}$	26.53	26.37	28.84	32.88	**34.47**	**35.70**	37.59	42.52	33.11
MF-AFNS	33.27	35.28	38.05	39.40	39.10	38.38	37.89	**36.97**	37.29
Panel C: Six-months-ahead forecasts									
MF-RNS$_{indep}$	64.93	68.72	72.13	72.80	71.72	68.85	67.63	69.07	69.48
MF-RNS$_{corr}$	**46.90**	**49.63**	**53.48**	**57.28**	57.69	56.36	57.06	62.23	**55.08**
MF-AFRNS	62.47	66.12	70.36	72.79	72.78	71.99	71.32	66.76	69.33
MF-DNS$_{indep}$	56.32	59.93	62.03	62.16	60.70	57.28	55.78	57.31	58.94
MF-DNS$_{corr}$	48.03	50.63	54.13	57.29	**57.38**	**55.92**	56.70	61.95	55.25
MF-AFNS	56.42	60.86	63.98	63.81	61.53	57.56	**55.13**	**50.44**	58.72
Panel D: One-year-ahead forecasts									
MF-RNS$_{indep}$	109.51	113.72	115.35	112.97	109.40	104.27	103.38	106.86	109.43
MF-RNS$_{corr}$	93.45	96.52	98.05	98.88	96.56	91.86	91.13	95.63	95.26
MF-AFRNS	105.21	109.84	112.58	112.41	110.41	108.14	107.59	102.52	108.59
MF-DNS$_{indep}$	**85.98**	**89.76**	**90.67**	**88.28**	**84.29**	**77.54**	**75.10**	77.34	**83.62**
MF-DNS$_{corr}$	94.61	97.38	98.48	98.94	96.47	91.63	90.78	95.10	95.42
MF-AFNS	88.94	93.29	94.95	92.38	87.84	80.66	76.16	**67.77**	85.25

Note: This table contains the Root Mean Squared Forecast Errors (RMSFEs) in basis points across maturities for all specifications and forecast horizons considered. The maturities are reported in months. The out-of-sample period encapsulates the period between January 2016 and February 2020. All specifications and forecasts are estimated according to Section 3. Bold entries indicate the cross-sectional lowest RMSFE for a respective horizon and maturity.

Table 12: RMSFEs (in basis points) from Maximum Likelihood estimation

	Maturity (Months)								
Model	**3**	**6**	**12**	**24**	**36**	**60**	**84**	**120**	**All**
Panel A: One-month-ahead forecasts									
RNS_{indep}	14.98	12.80	14.99	17.44	19.20	20.87	21.07	21.22	17.82
RNS_{corr}	18.96	14.25	13.26	16.57	18.66	20.77	21.46	22.11	18.26
AFRNS	15.63	12.41	14.73	17.55	19.28	20.95	21.29	23.25	18.14
DNS_{indep}	19.23	15.82	15.86	18.19	19.77	21.35	21.71	22.09	19.25
DNS_{corr}	18.48	13.80	13.08	16.48	18.59	20.63	21.20	21.75	18.00
AFNS	17.11	13.02	13.73	16.86	18.84	20.84	21.22	21.33	17.87
Panel B: One-quarter-ahead forecasts									
RNS_{indep}	23.94	25.72	29.80	32.55	34.47	35.94	35.79	35.33	31.69
RNS_{corr}	30.45	27.35	26.56	30.49	33.43	37.31	39.66	41.94	33.40
AFRNS	23.49	24.87	29.47	33.62	36.31	38.49	38.73	39.64	33.08
DNS_{indep}	33.42	32.17	32.52	35.49	37.36	38.96	39.35	39.86	36.14
DNS_{corr}	28.63	25.55	25.51	30.12	33.14	36.41	37.92	39.26	32.07
AFNS	25.59	24.79	27.50	32.39	35.64	38.78	39.50	38.96	32.89
Panel C: Six-months-ahead forecasts									
RNS_{indep}	41.26	45.94	51.83	55.16	56.45	55.61	53.17	50.07	51.19
RNS_{corr}	49.70	47.87	47.99	51.93	54.77	59.12	62.55	66.42	55.05
AFRNS	40.84	45.58	52.23	57.58	60.42	61.32	59.64	57.49	54.39
DNS_{indep}	55.54	55.18	55.99	59.63	61.67	62.98	63.09	63.42	59.69
DNS_{corr}	46.41	45.17	45.82	50.88	54.02	57.47	59.26	60.90	52.49
AFNS	43.27	44.80	48.16	53.66	57.47	60.97	61.37	59.61	53.66
Panel D: One-year-ahead forecasts									
RNS_{indep}	67.40	77.06	86.92	91.90	92.31	87.47	81.17	74.05	82.28
RNS_{corr}	84.84	84.69	84.06	87.20	89.12	93.56	99.11	106.90	91.19
AFRNS	66.95	76.43	86.24	91.73	92.60	88.47	82.62	76.23	82.66
DNS_{indep}	82.61	84.02	85.57	88.00	87.81	85.81	85.15	86.84	85.72
DNS_{corr}	78.10	78.68	79.81	84.79	86.81	88.73	90.61	93.72	85.16
AFNS	69.58	74.00	78.53	82.90	84.76	85.28	84.18	81.30	80.07

Note: This table contains the Root Mean Squared Forecast Errors (RMSFEs) in basis points across maturities for all specifications and forecast horizons considered. The maturities are reported in months. The out-of-sample period encapsulates the period between January 2016 and February 2020. All specifications and forecasts are estimated with classical methods according to Appendix C.3.

H.2 Figures

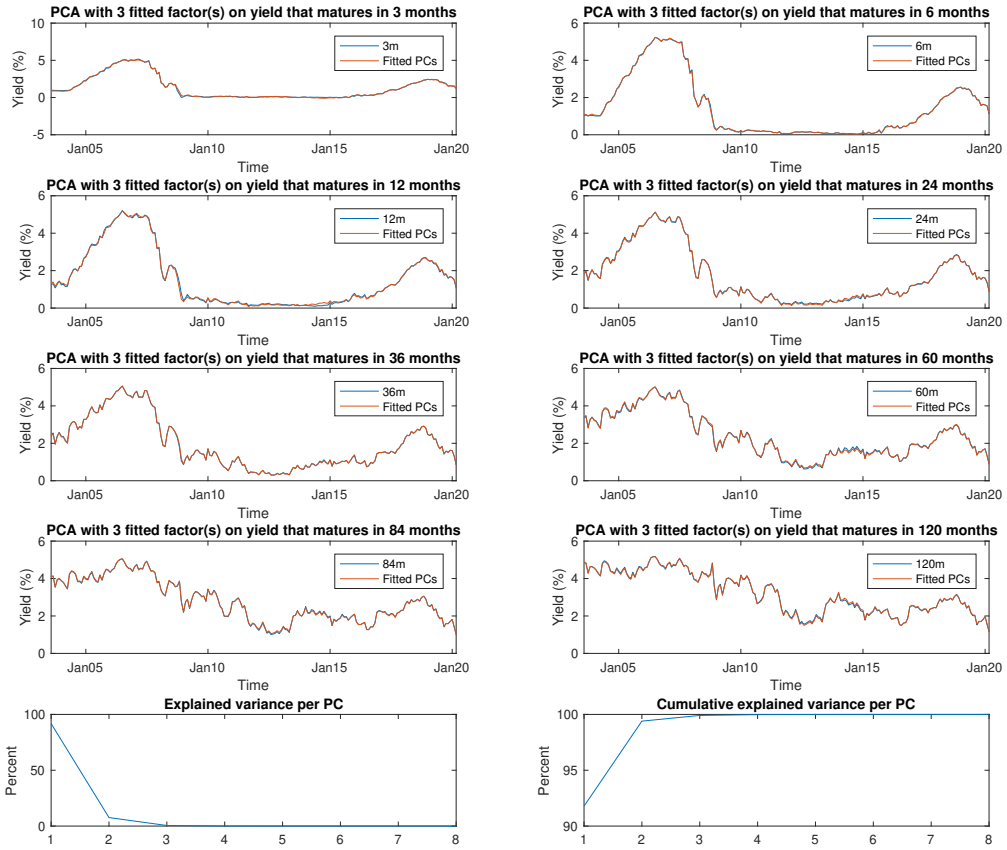

Figure 19: Principal Component Analysis (PCA) on the covariance matrix of the empirical U.S. data set of yields ranging from July 2003 to February 2020. The analysis lays the foundation and motivation for our derived specification, as it provides heuristic evidence of a three-factor state diffusion being sufficient to adequately capture the underlying term structure dynamics (Litterman (1991); Wright (2006)). The evidence is two-fold, i.e. the first four rows show the first three fitted Principal Components to follow the movements of the underlying yields. Secondly, the last row shows the individual and cumulative variance captured by the principal components where the first three principal components account to 99.9% of the underlying variance.

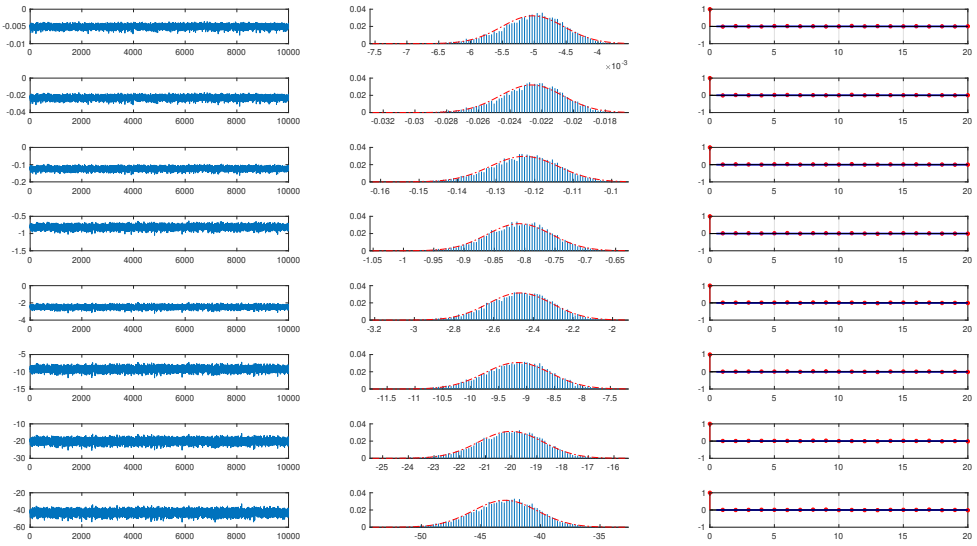

Figure 20: Traceplots, histograms with fitted posterior densities and Empirical Auto Correlation Functions (EACFs) for the yield-adjustment term in basis points of the AFRNS model. The order coincides with the vertical stacking of the components, i.e. $\boldsymbol{a_{3M}}$, $\boldsymbol{a_{6M}}$, $\boldsymbol{a_{1Y}}$, $\boldsymbol{a_{2M}}$, $\boldsymbol{a_{3Y}}$, $\boldsymbol{a_{5M}}$, $\boldsymbol{a_{7Y}}$, $\boldsymbol{a_{10Y}}$ from top to bottom (e.g. see Table 6).

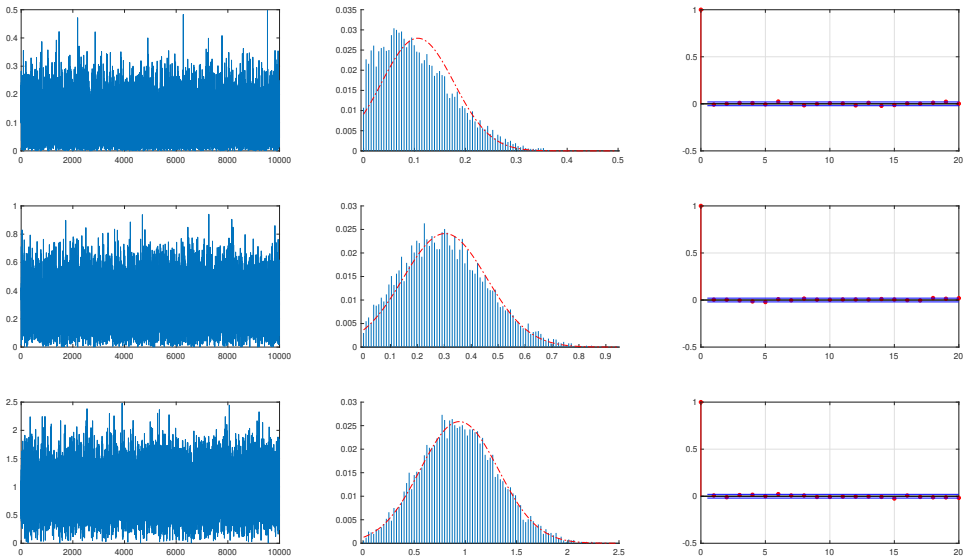

Figure 21: Traceplots, histograms with fitted posterior densities and Empirical Auto Correlation Functions (EACFs) for the risk-neutral state dynamics $\boldsymbol{K}^{\mathbb{P}}$ of the AFRNS model. The vertical order of the plots coincides with the diagonal organization of the risk-neutral dynamics in the second panel of Table 6.

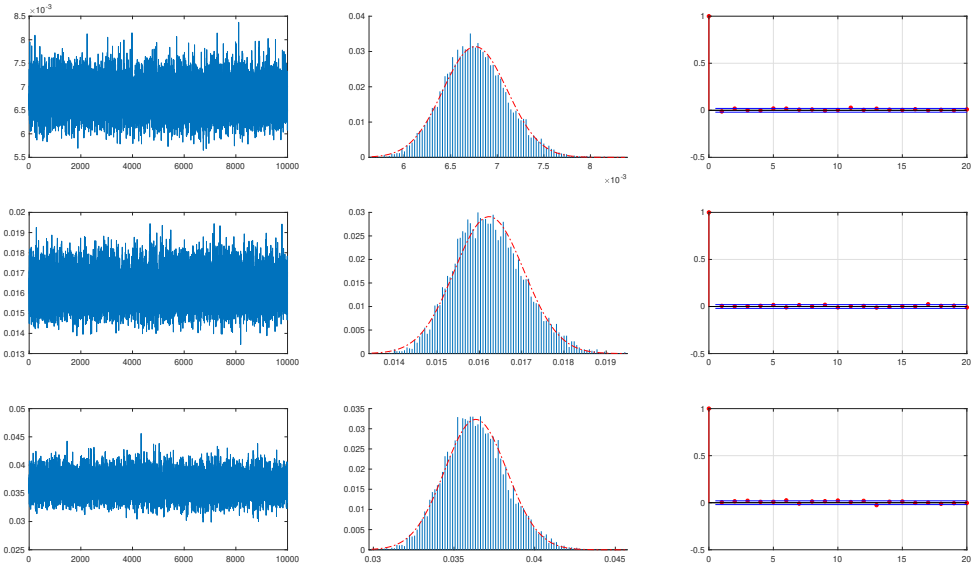

Figure 22: Traceplots, histograms with fitted posterior densities and Empirical Auto Correlation Functions (EACFs) for the risk-neutral volatilities $\mathbf{\Sigma}$ of the AFRNS model. The vertical order of the plots coincides with the diagonal organization of the risk-neutral volatilities in the second panel of Table 6.

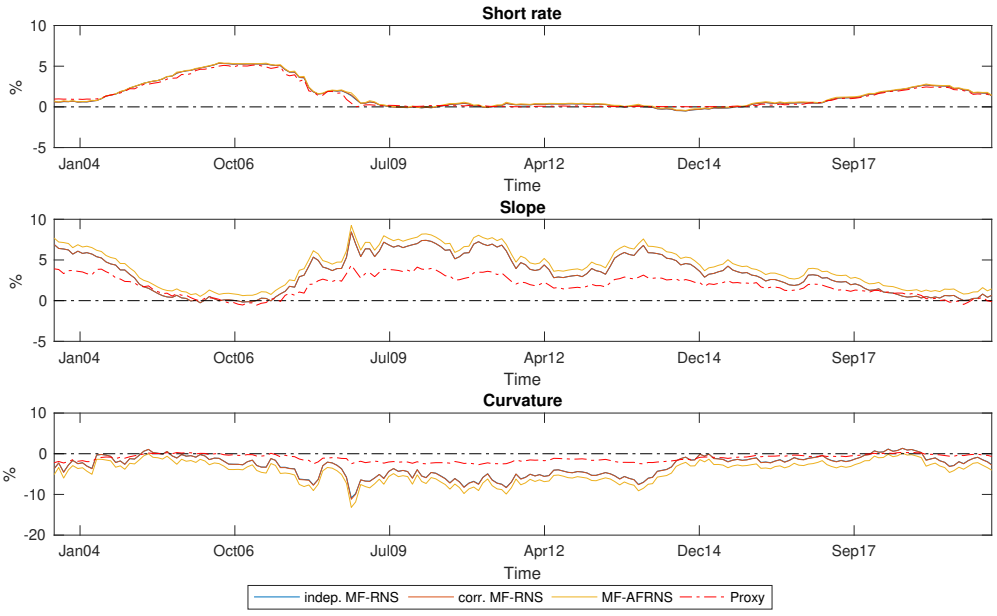

Figure 23: Time series of the latent factors from the MF-(AF)RNS models over the period July 2003 - February 2020. The red-dotted lines correspond to the respective empirical counterparts as mentioned in the second panel of Table 2.

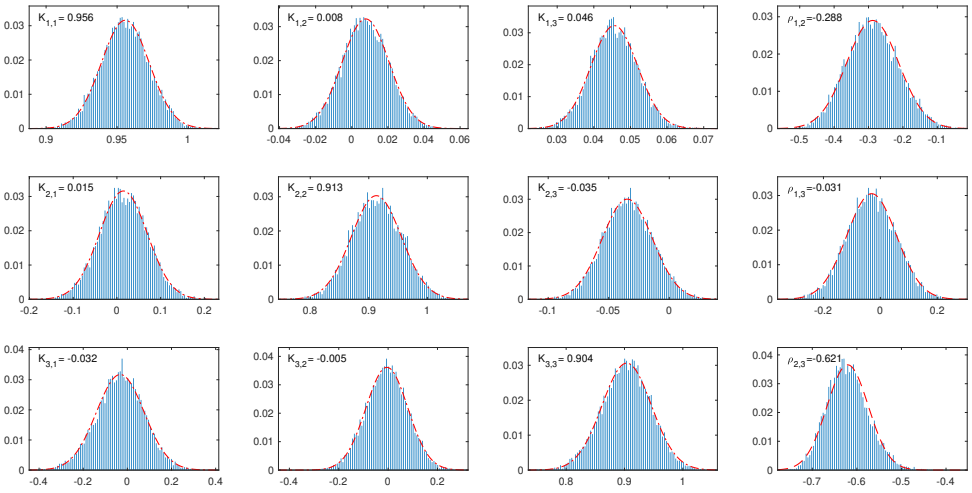

Figure 24: In-sample posterior densities of the state dynamics and disturbance correlations for the correlated factor MF-RNS model, where the values coincide with the posterior means. The red lines correspond to fitted normal distributions with means equal to the respective posterior means and variances equal to the squared respective posterior standard deviations for all parameters considered.

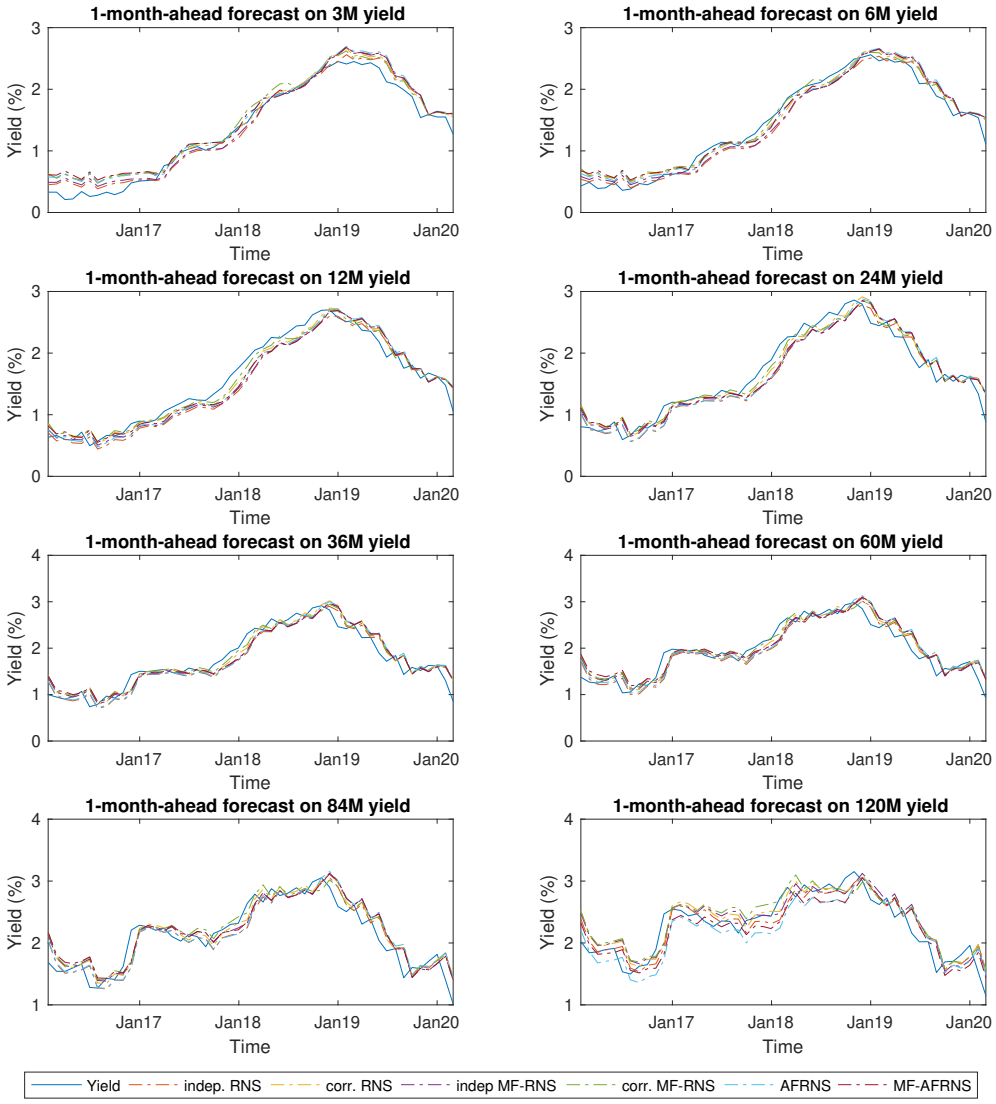

Figure 25: One-month-ahead forecasts for each maturity from January 2016 to February 2020. For each forecast, we re-estimate our model via a rolling window approach of 120 observations where we apply the same protocol as with our in-sample analysis to generate effective drawings. The forecast densities are generated according to step 8-9 in Section 3 where the point forecasts correspond to their respective posterior means.

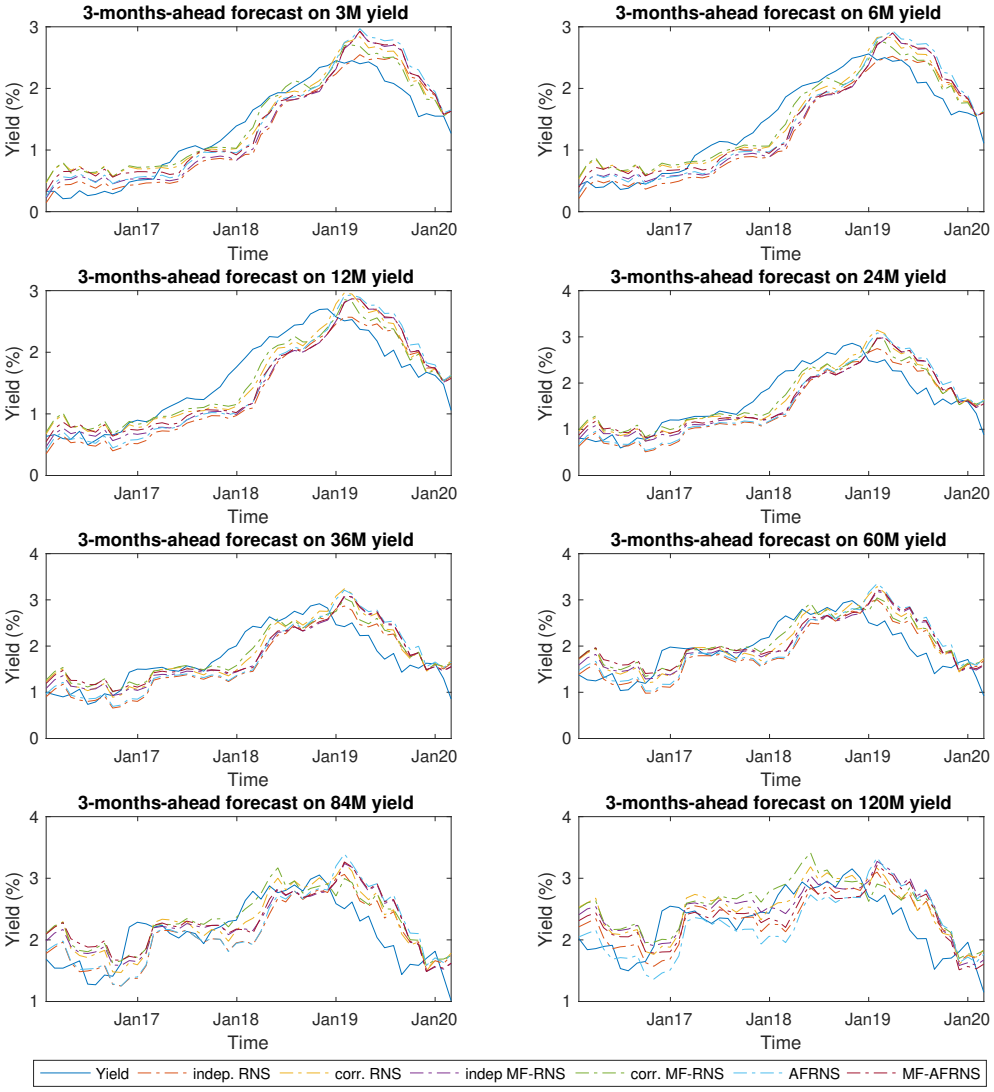

Figure 26: One-quarter-ahead forecasts for each maturity from January 2016 to February 2020. For each forecast, we re-estimate our model via a rolling window approach of 120 observations where we apply the same protocol as with our in-sample analysis to generate effective drawings. The forecast densities are generated according to step 8-9 in Section 3 where the point forecasts correspond to their respective posterior means.

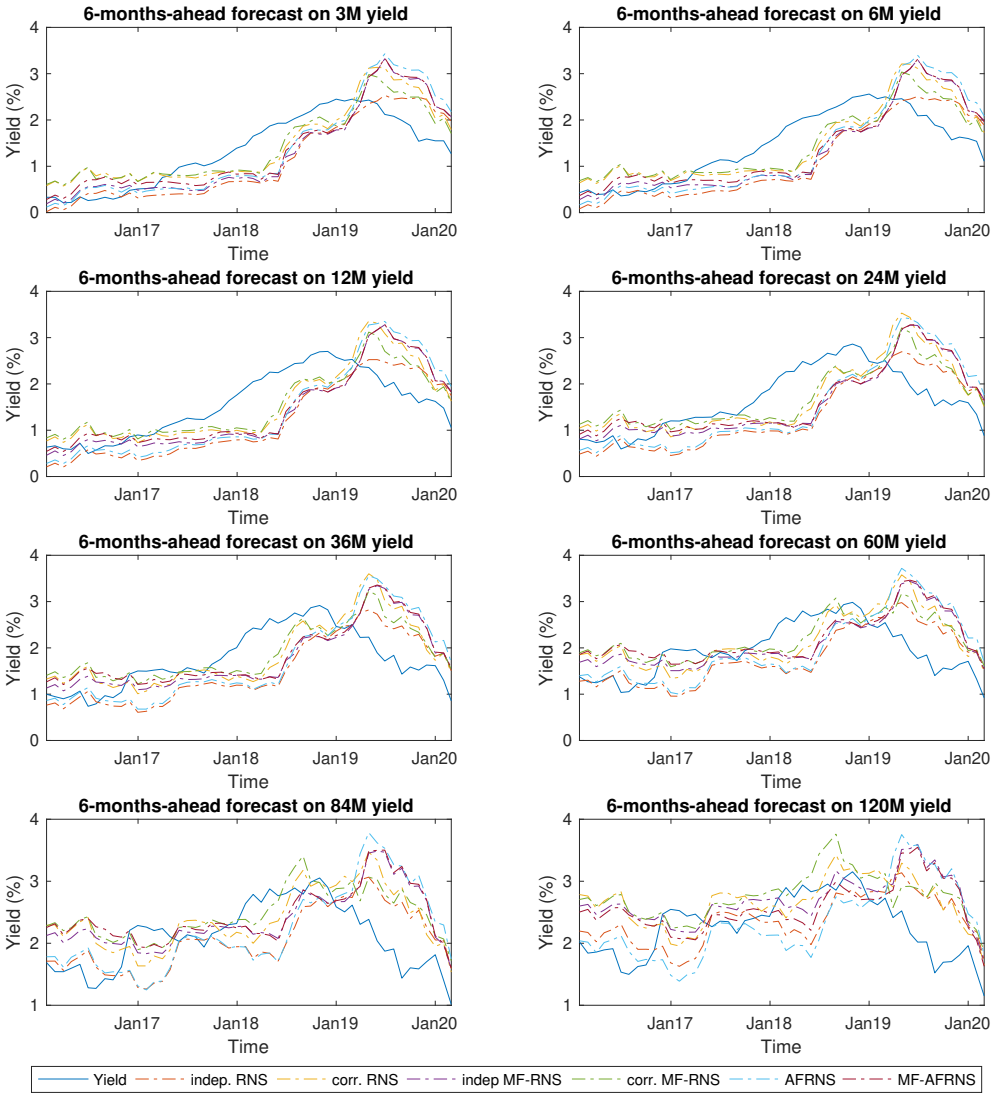

Figure 27: Six-months-ahead forecasts for each maturity from January 2016 to February 2020. For each forecast, we re-estimate our model via a rolling window approach of 120 observations where we apply the same protocol as with our in-sample analysis to generate effective drawings. The forecast densities are generated according to step 8-9 in Section 3 where the point forecasts correspond to their respective posterior means.

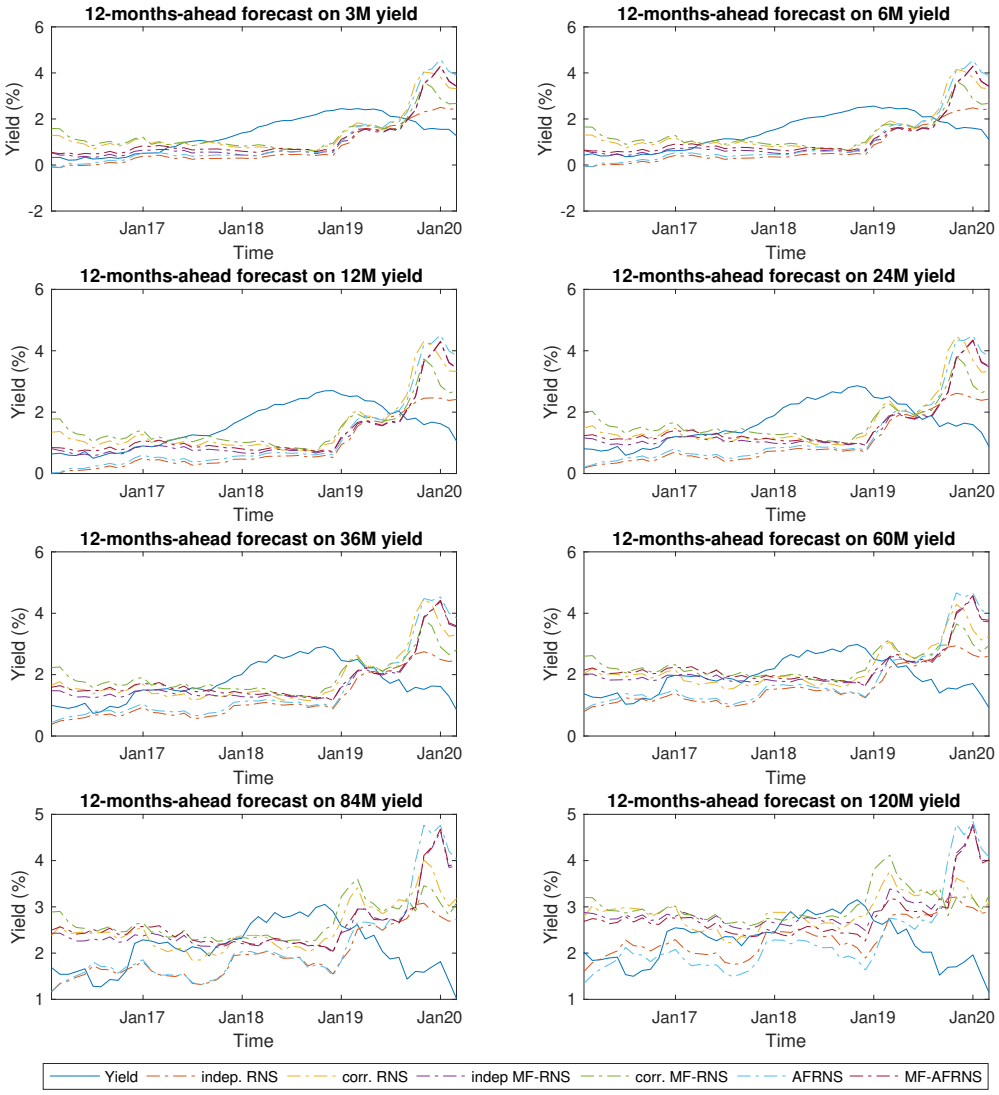

Figure 28: One-year-ahead forecasts for each maturity from January 2016 to February 2020. For each forecast, we re-estimate our model via a rolling window approach of 120 observations where we apply the same protocol as with our in-sample analysis to generate effective drawings. The forecast densities are generated according to step 8-9 in Section 3 where the point forecasts correspond to their respective posterior means.

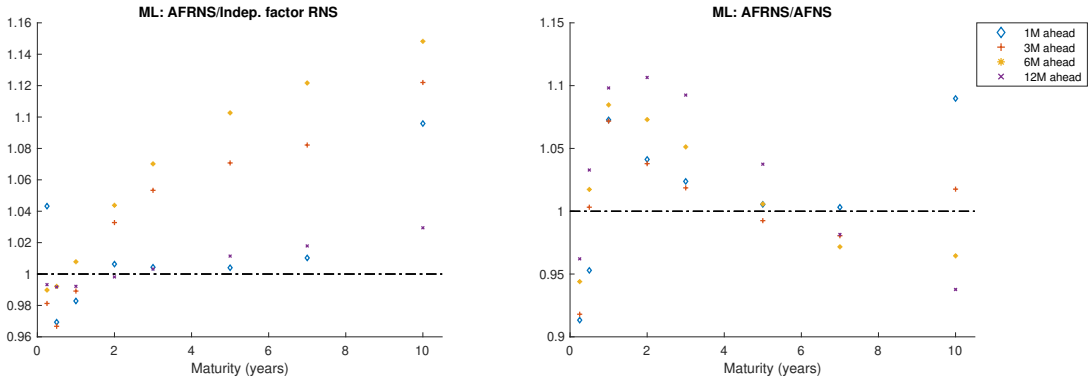

Figure 29: Relative RMSFEs to assess the predictive capability of the AFRNS model in a Maximum Likelihood (ML) environment. In the left partition, we provide the relative RMSFEs from the AFRNS model opposed to the independent factor RNS model. The right partition shows the relative RMSFEs of the AFRNS model compared to the AFNS model.

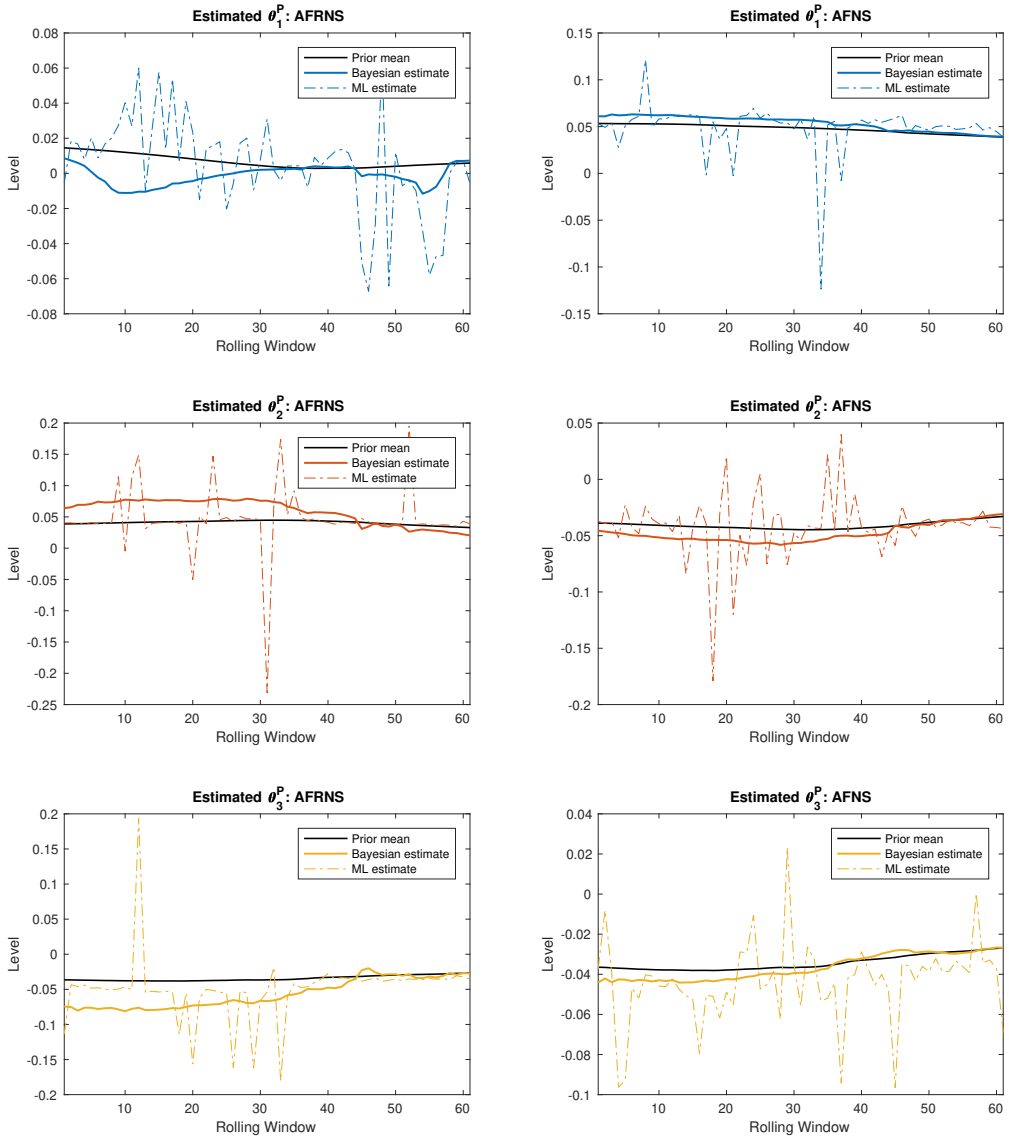

Figure 30: Estimates of the unconditional means from the AFNS model and the AFRNS model within our out-of-sample study. The black lines correspond to the $\mathcal{N}(\theta, 0.01)$ priors on the unconditional means from Diebold et al. (2008), the colored solid lines correspond to the posterior point estimates of the unconditional means generated via our Bayesian approach and the striped colored refer to the estimated unconditional means via classical estimation.